DECISION SUPPORT TOOL FOR PREDICTING AIRCRAFT
ARRIVAL RATES FROM WEATHER FORECASTS

by

David A. Smith
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Information Technology

Committee:

Dr. Lance Sherry, Dissertation Director
Dr. George Donohue, Committee Member
Dr. Rajesh Ganesan, Committee Member
Dr. Andrew Loerch, Committee Member
Dr. Aimee Flannery, Committee Member
Dr. Daniel Menasce, Associate Dean
for Graduate Studies and Research
Dr. Lloyd J. Griffiths, Dean, The Volgenau
School of Information Technology and
Engineering

Date: Spring Semester 2008
George Mason University
Fairfax, VA
Decision Support Tool for Predicting Aircraft Arrival Rates from Weather Forecasts

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

David A. Smith
Master of Science
Rensselaer Polytechnic Institute, 2000
Bachelor of Science
United States Military Academy, 1990

Director: Dr. Lance Sherry, Professor
Department of Systems Engineering and Operational Research

Spring Semester 2008
George Mason University
Fairfax, VA
Dedication

I dedicate this dissertation to my wife and best friend Jeanine, my buddy and son Gunner, and my princess and daughter Madeline.
Acknowledgments

I would like to thank my wife Jeanine for supporting me through an 18 year Army career. I have received many opportunities while she has taken care of the home and the children. Although I have written and been published, her work has led to the lasting legacy that is our wonderful children. I would also like to thank the United States Army for providing me with an outstanding military and civilian education. I only hope that I can live up to the expectations that come with this opportunity.

Two years ago I was in search of a topic for this dissertation. I went to a talk by Dr. David Rodenhuis which introduced me to aviation weather. Our early conversations led to this research and introduction to Dr. George Donohue, the Director of the Center for Air Transportation System Research (CATSR). Dr. Donohue introduced me to air traffic management and the problems facing the National Airspace System (NAS). We discussed potential topics and traded ideas on how to improve air traffic flow. He also introduced me to the Executive Director of the CATSR Lab, Dr. Lance Sherry. I owe almost everything I know about air traffic management to Dr. Sherry. He eventually agreed to be my dissertation director and has helped me focus my research to go beyond merely an academic document to one that could be used by the airline industry and the air traffic managers.

It is a pleasure to have Dr. Rajesh Ganesan, Dr. Andrew Loerch, and Dr. Aimee Flannery on my committee as well. I appreciate the time and effort they set aside to help me complete this dissertation.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Abstract</td>
<td>xiv</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Scope</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Research Problem</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Objectives</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Contribution</td>
<td>5</td>
</tr>
<tr>
<td>1.5.1 Uniqueness of Research</td>
<td>5</td>
</tr>
<tr>
<td>1.5.2 Benefits of Research</td>
<td>6</td>
</tr>
<tr>
<td>1.5.3 Structure of Dissertation</td>
<td>7</td>
</tr>
<tr>
<td>2 Review of Prior Research</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Airport Arrival Inefficiencies and Delay Reduction Strategies</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1 Airport Arrival Inefficiencies</td>
<td>8</td>
</tr>
<tr>
<td>2.1.2 Delay Reduction Strategies</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Reroute Around Weather</td>
<td>10</td>
</tr>
<tr>
<td>2.2.1 Dynamic Strategy Models</td>
<td>11</td>
</tr>
<tr>
<td>2.2.2 Controller Workload Models</td>
<td>12</td>
</tr>
<tr>
<td>2.2.3 Route Optimization Models</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Forecast Evaluation</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Weather Trends and Analysis</td>
<td>15</td>
</tr>
<tr>
<td>2.5 Delay Models</td>
<td>16</td>
</tr>
<tr>
<td>2.5.1 The Interactive Weather Brief</td>
<td>16</td>
</tr>
<tr>
<td>2.5.2 Surface Conditions and Runway Configurations</td>
<td>17</td>
</tr>
<tr>
<td>2.5.3 Regression Models</td>
<td>17</td>
</tr>
<tr>
<td>2.5.4 Bayesian Models</td>
<td>18</td>
</tr>
<tr>
<td>2.5.5 Histogram Method</td>
<td>19</td>
</tr>
<tr>
<td>2.5.6 Tree and Clustering Methods</td>
<td>19</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.6</td>
<td>Literature Review Summary</td>
</tr>
<tr>
<td>3</td>
<td>Methodology</td>
</tr>
<tr>
<td>3.1</td>
<td>Military Decision Making Process</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Military Planning</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Air Traffic Management Intelligence</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Planning Components</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Planning Tools</td>
</tr>
<tr>
<td>3.1.5</td>
<td>Decision Point Development</td>
</tr>
<tr>
<td>3.2</td>
<td>Weather Delay Prediction Tool</td>
</tr>
<tr>
<td>3.3</td>
<td>Data Sources</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Ground Delay Program</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Terminal Aerodrome Forecast</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Aircraft Arrival Rates</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Delay Data</td>
</tr>
<tr>
<td>3.4</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>3.5</td>
<td>Proposed Method</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Creating the TAF Vector</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Method</td>
</tr>
<tr>
<td>3.6</td>
<td>Alternative Methods</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Duality</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Nonlinear Hyperplane</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Modified Method</td>
</tr>
<tr>
<td>3.7</td>
<td>Performance Measures</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Evaluation of Binary Classifiers</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Receiver Operating Characteristic Curve</td>
</tr>
<tr>
<td>4</td>
<td>Philadelphia Case Study</td>
</tr>
<tr>
<td>4.1</td>
<td>Philadelphia International Airport</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Philadelphia Airport Layout</td>
</tr>
<tr>
<td>4.2</td>
<td>Linear Support Vector Machine Method</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Philadelphia Results</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Delay Prediction</td>
</tr>
<tr>
<td>4.3</td>
<td>Dual Formulation</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Dual Formulation Results</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Nonlinear Conclusions</td>
</tr>
<tr>
<td>4.4</td>
<td>Evaluation of Alternate Classification Methods</td>
</tr>
</tbody>
</table>
5.7 Dulles International Airport .............................................. 111
  5.7.1 Airport Layout ...................................................... 113
  5.7.2 Dulles International Airport Evaluation Method ............... 113
  5.7.3 Dulles International Results ..................................... 115
6 Strategies for the Delay Tool ............................................. 119
  6.1 Hub and Spoke Operations .......................................... 119
    6.1.1 Concept of the Reliever Hub .................................. 120
    6.1.2 Integrating the TAF Delay Tool into the Reliever Hub Concept ... 123
    6.1.3 Example Planning Day .......................................... 124
  6.2 Skip-Stop Operations ................................................ 125
    6.2.1 Southwest Airlines Operations ................................. 125
    6.2.2 Float Plane Operations ........................................ 126
    6.2.3 Example Float Plane Scenario .................................. 127
  6.3 Air Traffic Control System Command Center Planning ............. 128
    6.3.1 Mission Analysis .............................................. 128
    6.3.2 Decision Point Development .................................... 129
    6.3.3 Plan Refinement, Distribution, and Execution ............... 130
    6.3.4 Sample Planning Day .......................................... 131
7 Conclusions and Future Work ............................................ 133
  7.1 Weather Delay Prediction ............................................ 134
  7.2 Support Vector Machines ............................................ 135
    7.2.1 SVM Disadvantages ............................................. 135
    7.2.2 Future Work on SVMs .......................................... 136
  7.3 Air Traffic Management Planning ................................... 138
A Appendix A ............................................................... 139
B Appendix B ............................................................... 140
C Appendix C ............................................................... 141
D Appendix D ............................................................... 142
E Appendix E ............................................................... 143
F Appendix F ............................................................... 145
G Appendix G ............................................................... 146
H Appendix H ............................................................... 147
I Appendix I ............................................................... 156
  I.1 AAR 48 at 0800 ...................................................... 156
  I.2 AAR 52 at 0800 ...................................................... 162
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Macroscopic Capacity Model capacity analysis</td>
</tr>
<tr>
<td>3.1</td>
<td>Forecast Production Time and Cycle</td>
</tr>
<tr>
<td>3.2</td>
<td>Weather Forecast Codes</td>
</tr>
<tr>
<td>3.3</td>
<td>Demand hour and assumed time block</td>
</tr>
<tr>
<td>3.4</td>
<td>Possible Admissible Kernels</td>
</tr>
<tr>
<td>4.1</td>
<td>Philadelphia Training Data</td>
</tr>
<tr>
<td>4.2</td>
<td>Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.3</td>
<td>Tool Results for Philadelphia Training Data</td>
</tr>
<tr>
<td>4.4</td>
<td>Tool Results for Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.5</td>
<td>Philadelphia Delay Predictions</td>
</tr>
<tr>
<td>4.6</td>
<td>Predictions compared to training data</td>
</tr>
<tr>
<td>4.7</td>
<td>Predictions compared to testing data</td>
</tr>
<tr>
<td>4.8</td>
<td>Least Squares Philadelphia Training Data</td>
</tr>
<tr>
<td>4.9</td>
<td>Least Squares Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.10</td>
<td>Tool Results for Least Squares Philadelphia Training Data</td>
</tr>
<tr>
<td>4.11</td>
<td>Tool Results for Least Squares Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.12</td>
<td>Philadelphia Nearest Neighbor Testing Data</td>
</tr>
<tr>
<td>4.13</td>
<td>Tool Results for Nearest Neighbor Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.14</td>
<td>Recursive Partitioning Philadelphia Training Data</td>
</tr>
<tr>
<td>4.15</td>
<td>Recursive Partitioning Philadelphia Testing Data</td>
</tr>
<tr>
<td>4.16</td>
<td>Tool Results for Recursive Partitioning Philadelphia Training Data</td>
</tr>
<tr>
<td>4.17</td>
<td>Tool Results for Recursive Partitioning Philadelphia Testing Data</td>
</tr>
<tr>
<td>5.1</td>
<td>Newark Training Data</td>
</tr>
<tr>
<td>5.2</td>
<td>Newark Testing Data</td>
</tr>
<tr>
<td>5.3</td>
<td>Tool Results for Newark Training Data</td>
</tr>
<tr>
<td>5.4</td>
<td>Tool Results for Newark Testing Data</td>
</tr>
<tr>
<td>5.5</td>
<td>Newark Delay Predictions</td>
</tr>
<tr>
<td>5.6</td>
<td>O’Hare Training Data</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Air Traffic Control System Command Center Staff</td>
<td>23</td>
</tr>
<tr>
<td>3.2</td>
<td>Example TAF</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>Aircraft Arrival Demand Chart</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>IAD Hourly Demand</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Separating line defines a boundary</td>
<td>40</td>
</tr>
<tr>
<td>3.6</td>
<td>Full Separation requires a curve.</td>
<td>41</td>
</tr>
<tr>
<td>3.7</td>
<td>Objects are mapped using a set of mathematical functions.</td>
<td>42</td>
</tr>
<tr>
<td>3.8</td>
<td>Philadelphia Airport Map</td>
<td>44</td>
</tr>
<tr>
<td>3.9</td>
<td>ROC Graph Example</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Philadelphia Airport Map</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>PHL Hourly Demand</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Weather Channel Airport Impact Map</td>
<td>62</td>
</tr>
<tr>
<td>4.4</td>
<td>PHL Training Data Kernel ROC Graph</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>PHL Testing Data Kernel ROC Graph</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>Training Data ROC Graph</td>
<td>77</td>
</tr>
<tr>
<td>4.7</td>
<td>Testing Data ROC Graph</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Newark Airport Map</td>
<td>80</td>
</tr>
<tr>
<td>5.2</td>
<td>EWR Hourly Demand</td>
<td>81</td>
</tr>
<tr>
<td>5.3</td>
<td>Chicago O'Hare Airport Map</td>
<td>86</td>
</tr>
<tr>
<td>5.4</td>
<td>ORD Hourly Demand</td>
<td>87</td>
</tr>
<tr>
<td>5.5</td>
<td>Atlanta Airport Map</td>
<td>92</td>
</tr>
<tr>
<td>5.6</td>
<td>ATL Hourly Demand</td>
<td>93</td>
</tr>
<tr>
<td>5.7</td>
<td>ATL AAR Frequency</td>
<td>94</td>
</tr>
<tr>
<td>5.8</td>
<td>LaGuardia Airport Map</td>
<td>99</td>
</tr>
<tr>
<td>5.9</td>
<td>LGA Hourly Demand</td>
<td>100</td>
</tr>
<tr>
<td>5.10</td>
<td>John F. Kennedy International Airport Map</td>
<td>104</td>
</tr>
<tr>
<td>5.11</td>
<td>JFK Hourly Demand</td>
<td>105</td>
</tr>
</tbody>
</table>
Abstract

DECISION SUPPORT TOOL FOR PREDICTING AIRCRAFT ARRIVAL RATES FROM WEATHER FORECASTS

David A. Smith, PhD
George Mason University, 2008
Dissertation Director: Dr. Lance Sherry

Over the years airline delays have worsened so much so that the problem is attracting Presidential attention. During the 2007 Thanksgiving and Christmas travel season, the President attempted to alleviate these delays by opening up military airspace to civilian transport. Although this may have provided some relief it did not address the issue that the principle “bottlenecks” of the air traffic control system are the major commercial airports. Atlanta, Detroit, St. Louis, Minneapolis, Newark, Philadelphia, and LaGuardia all expect to be at least 98% capacity by 2012. Due to their cost and the environmental and noise issues associated with construction, it is unlikely that any new airports will be built in the near future. Therefore to make the National Airspace System run more efficiently, techniques to more effectively use the limited airport capacity must be developed.

Air Traffic Management has always been a tactical exercise, with decisions being made to counter near term problems. Since decisions are made quickly, limited time is available to plan out alternate options that may better alleviate arrival flow problems at airports. Extra time means nothing when there is no way to anticipate future operations, therefore predictive tools are required to provide advance notice of future air traffic delays. This research provides two essentials needed to more efficiently use the limited airport capacity.
First, it introduces the Military Decision Making Process (MDMP) to civilian air traffic management. The MDMP is an established and proven analytical process that assists the commander and staff in developing estimates and a plan. This process can be modified for civilian use in order to simplify and standardize the planning process in order to develop options to handle potential problems. Second, this research describes how to use Support Vector Machines (SVM) to predict future airport capacity. The Terminal Aerodrome Forecast (TAF) is used as an independent variable within the SVM to predict Aircraft Arrival Rates (AAR) which depict airport capacity. Within a decision support tool, the AAR can be derived to determine Ground Delay Program (GDP) program rate and duration and passenger delay.

The research compares the SVM to other classification methods and confirms that it is an effective way to predict airport capacity. The research goes on further to describe how to integrate the SVM method into a decision support tool and then use that tool within the MDMP that has been modified for civilian air traffic management. Real world examples are included to highlight the usefulness of this research to airlines, air traffic managers, and the flying consumer. New strategies to minimize the effect of weather on arrival flow are developed and current techniques are discussed and integrated into the process. The introduction of this decision support tool will expand the amount of time available to make decisions and move resources to implement plans.
Chapter 1: Introduction

1.1 Background

Air traffic congestion has become a widespread phenomenon in the United States. The principle bottlenecks of the air traffic control system are the major commercial airports, of which at least a dozen currently operate near or above their point of saturation under even moderately adverse weather conditions [1]. The Macroscopic Capacity Model (MCM) analyzed 16 airports within a 1000 nmi. triangle from Boston, Massachusetts, to Minneapolis, Minnesota, to Tallahassee, Florida. Based on this analysis, the MCM showed that in 1997 these airports were operating at 74% of maximum capacity. The model further went on to predict that these airports will be at 89% capacity by 2012 [2]. Table 1.1 shows the individual capacities for each of the 16 airports for 1997 and the predictions for 2012. At this time there are no new major airports planned. Currently, several government and industry efforts are underway in the United States and Europe to examine ways to alleviate the pressure on capacity-constrained airports. Building additional runways is not feasible as the sole remedy to this problem, due to high costs, construction time, space limitations, and environmental concerns [3]. Therefore, it is important that new solutions are developed to use the available airport capacity more efficiently.

1.2 Scope

Efficiency is measured in different time horizons and different levels of detail. In this dissertation efficiency is defined as maximizing all available airplane landing capacity at an airport. Airport capacity can be measured in many levels, but in this dissertation the unit of measure of capacity is Aircraft Arrival Rates (AAR). Aircraft Arrival Rate in the number
Table 1.1: Macroscopic Capacity Model capacity analysis

<table>
<thead>
<tr>
<th>Airport</th>
<th>1997</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago (ORD)</td>
<td>73%</td>
<td>91%</td>
</tr>
<tr>
<td>Atlanta (ATL)</td>
<td>78%</td>
<td>98%</td>
</tr>
<tr>
<td>Detroit (DTW)</td>
<td>77%</td>
<td>99%</td>
</tr>
<tr>
<td>St. Louis (STL)</td>
<td>80%</td>
<td>99%</td>
</tr>
<tr>
<td>Minneapolis (MSP)</td>
<td>93%</td>
<td>100%</td>
</tr>
<tr>
<td>Charlotte (CLT)</td>
<td>73%</td>
<td>95%</td>
</tr>
<tr>
<td>Boston (BOS)</td>
<td>77%</td>
<td>86%</td>
</tr>
<tr>
<td>Newark (EWR)</td>
<td>94%</td>
<td>99%</td>
</tr>
<tr>
<td>Pittsburgh (PIT)</td>
<td>53%</td>
<td>69%</td>
</tr>
<tr>
<td>Philadelphia (PHL)</td>
<td>91%</td>
<td>99%</td>
</tr>
<tr>
<td>Cincinnati (CVG)</td>
<td>55%</td>
<td>95%</td>
</tr>
<tr>
<td>New York (JFK)</td>
<td>70%</td>
<td>80%</td>
</tr>
<tr>
<td>LaGuardia (LGA)</td>
<td>96%</td>
<td>99%</td>
</tr>
<tr>
<td>Washington Dulles (IAD)</td>
<td>47%</td>
<td>61%</td>
</tr>
<tr>
<td>Washington Reagan (DCA)</td>
<td>66%</td>
<td>70%</td>
</tr>
<tr>
<td>Cleveland (CLE)</td>
<td>63%</td>
<td>87%</td>
</tr>
</tbody>
</table>

of aircraft that can be landed at an airport per hour. The optimal AARs are published by the FAA in the Airport Capacity Benchmark Report. The FAA has developed capacity benchmarks for 35 of the nation’s busiest airports. Capacity benchmarks are defined as the maximum number of flights an airport can routinely handle in an hour, for the most commonly used runway configuration in each specified weather condition. These benchmarks are estimates of a complex quality that varies widely with weather, runway configuration, and the mix of aircraft types. Capacity benchmarks assume there are no constraints in the en route system or the terminal area [4]. The maximum airport AAR is the optimal benchmark which typically occurs during periods of unlimited ceiling and visibility. When the weather deteriorates, controllers increase the separation between aircraft and decrease the AAR.

When the system is congested, the FAA slows down the flow of arrivals into an affected airport by imposing a GDP at appropriate departure airports. Typically, a GDP occurs when inclement weather causes controllers to increase separation space within the terminal
area. When the FAA imposes a GDP, it uses the Enhanced Traffic Management System (ETMS) to forecast traffic patterns and volume. ETMS gives each affected aircraft an expected departure clearance time (EDCT), and these EDCTs can change as conditions improve or demand decreases. Within the affected geographical areas, EDCT “slots” are initially assigned based on the flight plan’s schedule.

Once the FAA institutes the GDP, the airlines can then employ two different strategies. First, at any time during a GDP, an individual airline can rearrange its own slot assignments, or substitute, to optimize its operations. This is one of the most important features of the FAA’s Collaborative Decision Making (CDM) program, which was instituted in the 1990s partly in an effort to decentralize air traffic control decisions that have an economic impact on airlines. Second, the Slot Credit Substitution (SCS) program allows for the anonymous transfer of slots between airlines on a limited basis namely, a one-to-one slot bartering arrangement driven by a situation in which a carrier cannot use its departure slot [5].

The intent of this dissertation is not increase the airport capacity during optimal conditions, but instead predict when conditions are less than optimal and then provide a planning framework to allow air traffic managers and airlines to employ strategies to reduce delay in the system. The Ground Delay Program (GDP), connects the FAA to the airlines; the FAA inacts a GDP while the airlines react to it.

1.3 Research Problem

The congestion problem is made worse because most airline schedules are optimized without any consideration for unexpected irregularities. When irregularities occur, the primary goal of the airlines is to get back to the original schedule as soon as possible, while minimizing flight cancellations and delays [6]. When trying to get back on schedule, sometimes it is the complexity of the situation, coupled with time pressure, which results in results in quick decisions that may be less than optimal [7]. Therefore, it would be advantageous to develop techniques to lessen the complexity of the situation and increase the time available. The
military deals with complex situations similar in scope to the present day NAS by using a systematic planning process that can be adapted to air traffic management. The military decision-making process (MDMP) is a single, established, and proven analytical process that assists the commander and staff in developing estimates and a plan [8].

Despite the widespread belief that operations in complex, dynamic domains engage in planning, research with air traffic controllers has usually characterized the activities of the air traffic controller as largely tactical in nature [9]. Although controllers do come up with solutions, lack of planning leads to decisions that are inconsistent and unpredictable. Recent studies have demonstrated evidence of more strategic planning behavior if the controller was given sufficient time [10]. Consistency and predictability will allow stakeholders more time to respond to airspace restrictions and help reduce the delay to consumers within the system. Planning the National Airspace System (NAS) day may provide such consistency and with the plan briefed to NAS stakeholders, provide a chance to react to potential delays.

Since the MDMP is an already established procedure, only modest modifications are required to adapt the process to civilian air traffic process. MDMP also requires tools to produce operational, personnel, mechanical, and weather estimates for the coming day. Much research has been done and some tools are now available, but only require a systematic method for integrating the tool’s output into the decision making process. The MDMP provides this method.

A tool that is missing is one to produce an estimate of the impact of weather on inbound flight operations. Weather reports such as the Terminal Aerodrome Forecast (TAF), Aviation Routine Weather Report (METAR), and the Collaborative Convective Forecast Product (CCFP) all provide raw weather forecast information. None of these forecasts though inform NAS stakeholders what the effect of that weather will be on flight operations. This research intends to fill this void by developing a process from which a forecast can be entered to produce estimate of the delay and capacity of the airport within the forecast area. Capacity estimates, in the form of Aircraft Arrival Rates (AAR) are produced for four time periods of the operational day. Ground Delay Program estimates of duration
and program AARs along with expected delays can be derived from the predicted AARs. Now the forecast will not only provide the winds and ceiling, but also the AARs, GDPs, and expected delay.

1.4 Objectives

- Create a set of procedures and rules to plan the day’s operations, predict potential irregular operations, and create plans to counter any restrictions to flow in order to reestablish the original schedule as soon as possible.
- Create a Support Vector Machine algorithm to predict Aircraft Arrival Rates (AAR).
- Test and compare the different kernel methods for SVM and other classification methods.
- Use the Support Vector Machine algorithm in a method that predicts aircraft arrival rates at a given airport for given time periods of the day.
- Insert the method into a decision support tool that inputs the daily TAF and predicts GDPs, average delay at given time periods, and aircraft arrival rates.

1.5 Contribution

1.5.1 Uniqueness of Research

Previous research in this field has either focused on weather prediction or methods to decrease delay caused by weather. This research will use weather forecasts to predict reduced airport capacity and avoid delays. This research is unique in the following ways:

- The MDMP standardizes the planning for air traffic managers and all subordinate stakeholders.
- The research uses a classification method to expand the weather forecast beyond meteorological terms and into operational effects on an airport.
• Estimates delays using historical data instead of subjective interviews with controllers [11].

1.5.2 Benefits of Research

There are several potential benefits of this research.

• For the Air Traffic Control System Command Center (ATCSCC), adapting the Military Decision Making Process to air traffic management will produce a systematic and consistent procedure for planning an operational day in the NAS. Since all personnel will know and understand the process, then if a member of the air traffic management team is absent, then another member will be able to fill in at the position with limited loss of institutional knowledge.

• Both the ATCSCC and the airlines can use a tool that inputs the current Terminal Aerodrome Forecast (TAF) and produces airport capacity predictions hours in advance. There are two major barriers to air traffic flow, schedule congestion and weather. Schedule congestion can be predicted by using simulation tools that can fly the current commercial airline schedule. What is lacking though is a way to link weather with reduced airport capacity. This is done by connecting the TAF to AARs.

• The prediction tool helps with the weather part by providing the time and place of weather delays. Given this information, air traffic managers can create branch plans to counter any potential delays. These branch plans can include instituting GDPs at affected airports, implementing miles-in-trail restrictions into congested sectors, or in some cases do nothing but make all stakeholders aware of potential obstacles to flow. They can then publish these plans so the airlines know what may happen and what actions are going to be taken if congestion does occur. This increased predictability should increase NAS efficiency.

• Delays in the NAS are often caused by the actions of air traffic managers who are reacting to potential poor weather conditions that could compromise safety. This
research gives insight to how controllers react to forecasts and can further be used in simulations that require operational effects of weather.

1.5.3 Structure of Dissertation

This dissertation proposal is organized as the follows: Chapter 2 examines the relevant research in four areas:

- Reroute around weather research,
- Literature on forecast evaluation,
- Weather trends and analysis,
- Literature on delay models.

Chapter 3 describes the methodology which includes a description of the data and a brief introduction of support vector machines. Chapter 4 is a case study using Philadelphia International Airport as a test case. All forms of the SVM are tested as well as other classification methods. Chapter 5 reviews the SVM results for the other airports in the study. Chapter 6 describes some of the uses of the Weather Delay Prediction Model and Chapter 7 summarizes the conclusions and future work.
Chapter 2: Review of Prior Research

This dissertation is concerned with predicting airport delays using terminal aerodrome forecasts. This chapter examines the relevant research in four areas:

- Airport arrival inefficiencies and delay reduction strategies,
- Reroute around weather research,
- Literature on forecast evaluation,
- Weather trends and analysis,
- Literature on delay models.

Each area is discussed below.

2.1 Airport Arrival Inefficiencies and Delay Reduction Strategies

As discussed earlier, the chokepoints in the National Airspace System (NAS) exist at the airports. These delays are made worse by the inefficient use of available terminal airspace and landing slots. This research focuses on indentifying these inefficiencies and proposing solution to improve the flow.

2.1.1 Airport Arrival Inefficiencies

Robyn [5] summarizes some of the inefficiencies that exist within the system. One of the points made in the paper was that the FAA is offering the wrong incentives. One example is the tax financing of the air traffic control system. Tax financing encourages
overuse of scarce ATC capacity in part because commercial airlines pay for that capacity only indirectly, through passenger taxes, rather than directly. Moreover, a small aircraft contributes less in taxes than a large one, even though it costs the system about the same to serve them. For example, a 140-seat Airbus A320 flying from Denver to Phoenix contributes $1498 in taxes whereas a 50-seat regional jet pays only $502.

Robyn further explains that the incentive for the airline is now to substitute two smaller jets for one large jet to offer more frequent flights which is a major draw to business passengers. Ironically, this airline dysfunctional norm actually allows a solution to reduce the demand at an airport with reduced arrival capacity. If airport arrival capacity is reduced, then an airline can reduce the number of flights by substituting a larger jet to reduce arrival overall arrival demand.

The second inefficiency Robyn discussed was separation. Although, it was not explained fully in the paper, moving separation from the responsibility of the air traffic controller to the aircraft pilot would make more efficient use of available airspace. Obviously, decreasing separation space would increase the capacity of the airspace, but controllers may be uncomfortable with such close spacing. During inclement weather separation spacing is increased by controllers to insure safety.

These methods described in this dissertation do not allow for decreased spacing during optimal conditions, however, the methods may provide insight to how and why controllers reduce separation in response to weather forecasts. This may show instances where increased separation was unnecessary and therefore induced delays that could of been avoided.

2.1.2 Delay Reduction Strategies

Meyer et al. [12] highlights the affect of inclement weather that forces instrument flight rules (IFR) as opposed to visual flight rules (VFR) which are in effect during optimal weather conditions. There paper states that as a result of IFR separation rules, airport capacity under IFR is typically about 50% lower than its VFR capacity, due to the separation requirements. Under VFR, with 1.5 nautical mile spacing, a single runway can handle
a total of at least 60 arriving or departing aircraft per hour. Under IFR, arrivals and departures must be sequentially spaced 3 nautical miles, or approximately two minutes apart.

With a 50-50 mix of alternating arriving and departing aircraft, a single runway system is unaffected by operation under IFR. However, a closely spaced parallel runway system is severely affected by inclement weather. Since air traffic controllers become responsible for traffic separation under IFR, they must rely on radar to enforce adequate spacing. Current radar procedures cannot ensure proper lateral separation when parallel runways are spaced less than 4000 feet apart. As a result, closely spaced parallel runways are known as dependent parallels and are considered as a single runway system under IFR. Dependent parallels together can handle a maximum of 30 arrivals and 30 departures per hour. Under VFR, both runways together can handle at least 60 arrivals and 60 departures per hour.

The greatest strength of Meyer et al. is that it offers a solution to avoiding weather delays at hub airports. The paper recommends implementing a set of reliever-hubs to alleviate delays caused by weather. The best location for reliever-hubs must be carefully chosen to minimize all associated costs. Costs derived for potential weather delays, additional fuel consumption, traffic congestion, and infrastructure investment will be used in a mixed integer programming model to find the ideal hub locations.

This alternate hub strategy coupled with currently used airline strategy to counter GDP can be used relieve delay. No with the algorithm and model descibed in this dissertation more time is created to come up with a more optimal combination of solutions. Also, now there is time to allocate resources to needed locations to take advantage of these new procedures.

2.2 Reroute Around Weather

Much of the research associated with weather delays is concerned with en route travel. This research is of a tactical nature and involves either algorithms designed to efficiently reroute traffic using basic geometry or procedures that reduce the congestion in affected sectors.
2.2.1 Dynamic Strategy Models

Dynamic strategy models rely on building alternate routes based on step data that is either time based or position based. Each step provides a finite set of potential directions where the algorithm determines the optimal direction.

Krozel et al. [13] reviews variations on the Standard Arrival Routes (STAR) algorithm. Navigational aids of a potential arrival route to an airport are varied by a set amount of lateral separation to design a new arrival route as a function of time that safely avoid hazardous weather.

Nilim et al. [14] provides a dynamic routing strategy that minimizes the expected delay of the overall system while satisfying the consideration of the constraints obtained by the sector capacity as well as avoiding conflicts among the aircraft. Nilim also claims to have a generic weather model where the predicted zone has more than two different states.

For the weather uncertainty model, various weather teams produce predictions that some zones of airspace will be unusable for a certain time interval. Instead of assuming the outcomes of the storm as binary states which assume the best or worst case, the model attempts to simulate intermediate forms by using an \( n \)-state Markov chain with discrete changes at 15 minute intervals.

Actions of the Markov Decision Process are the directions of all \( N \) aircraft to follow in each stage with different realizations of weather. Options for potential aircraft directions are determined by calculating the points that can be reached in the next 15 minutes given their current positions. The paper does not further elaborate and how each re-route is determined and there is no discussion to explain the geometry required to determine the distances.

Davidson et al. [15] tries something similar except utilizes decision trees to model weather uncertainty. Probabilistic decision trees are first used to characterize potential weather outcomes and then to identify candidate TFM actions. The final picture is a set of decision trees that combines both types of information. This requires defining the relationships between probabilistic decision trees, ATM weather avoidance practices, TFM
decision points, and uncertain weather forecasts.

This paper calls for the development of algorithms that would define the relationships between the Probabilistic Decision Tree, weather avoidance practices, TFM decision points and uncertain weather forecasts. The concept here is similar to most dynamic programming approaches where they propose to change the potential options as conditions change, unfortunately how do we find these potential outcomes and their associated probabilities?

2.2.2 Controller Workload Models

Controller workload models are concerned with not overloading the controllers within the ATC system. They are focused on workload metrics and develop algorithms to ensure sectors do not exceed workload capacity.

Evans [16] argues that avoiding excessive controller workload situations that might result in aircraft separation violation is a very important consideration. The paper argues that controllers need better access to timely, high quality storm locations and severity data and complains that the coordination of route changes can be difficult because of the need to coordinate across sector boundaries. The paper offers more problems then solutions.

Brennan et al. [17] measures sector capacity based on controller workload. The basis of the proposal is in three parts:

- Plan a customer-preferred path to a decision point some distance away from the forecasted area.
- Develop a flight plan for a reroute around the forecasted area.
- Allow flights to traverse the weather area if able and if Air Traffic Control (ATC) workloads permit.

When action is required, the Air Traffic Control System Command Center (ATCSCC) uses the Flow Evaluation Area (FEA) capability to construct a thin FEA along an arc some distance from the predicted weather decision area. The FAA identifies flights within the
decision area and identifies navaids or fixes. Constraint Avoidance Routes to Destination (CARDs) are ATC preferred routes that avoid the forecasted convective weather and may join ATC-preferred routes to destination. Flights predicted to traverse the decision area file customer preferred routings to a decision point of their choice then along a CARD to their destination. If weather does not develop, flights proceed as filed to a decision point and then along a CARD to destination. If and only if controller workload permits, some flights may be routed through the area where severe weather has been forecasted, thus possible receiving a customer-preferred routing from departure to destination.

This process relies on the knowledge of how much work can be handled by controllers. One of the difficulties of this method is determining how much work can be handled by controllers with different experience in different sectors.

### 2.2.3 Route Optimization Models

Route optimization models take a given set of known routes and determines the optimal route. Krozel et al. [18] examined the Flow-Based Route Planner (FBRP) and Free Flight.

In the (FBRP), multiple, non-intersecting routes are designed to lead from the 200 nmi range ring to the airport metered fixes. Multiple routes are synthesized for each metering fix, based on maximizing the total number of routes that both avoid hazardous weather and do not cross over each other. Given a set of routes leading to the metering fixes, the route that minimizes the distance traveled, subject to the various constraints, is chosen for each aircraft. Aircraft are metered, using speed control, so that only one aircraft arrives at the metering fix at a time, regardless of the weather avoidance routes chosen.

In Free Flight, routes are synthesized such that the routes do not cross over hazardous weather and do not create separation conflicts with any aircraft already ahead of an aircraft in the transition airspace. The algorithm is “greedy” in the sense that each aircraft acts independently, selecting the best safe path between the 200 nmi range ring and the metering fix, subject to the constraints imposed by earlier aircraft arriving to the airspace.

It is often the proximity of weather to key resources that limits the throughput, and
therefore overall capacity. This is true for today's existing resources and was found to be true for conditions within all of the routing algorithms that were included in this study (e.g. entry points and metering fixes). Making these resources flexible and available is the key to maintaining throughput and increasing capacity during severe weather events.

2.3 Forecast Evaluation

Refining the forecast and forecast techniques are important to weather forecast professionals, therefore much of the literature focused on forecast accuracy and improvement. Because it is a relatively new product, the Collaborative Convection Forecast Product (CCFP) receives most of the attention.

Torbert and Rodenhuis [19] evaluated the effectiveness of the CCFP. This report was designed to show forecasters the impact of the CCFP on air flow management, either good or bad. It also describes how accurate the forecasts were with respect to location and timing and how consistent the forecasts were from one issuance to another.

Fahey and Rodenhuis [20] reviewed the performance of the CCFP, accessed it’s value and recommended future actions to increase it’s usefulness to air traffic management. The paper also concluded that before aviation weather becomes probabilistic, we must determine how probabilistic forecasts should be applied to deterministic strategic traffic management and tactical adjustment.

Kay et al. [21] concludes that CCFP forecasts are defined to be valid at a snapshot in time rather than valid over a certain time window. Also, for future CCFP definitions, a clarification of the time period over which the movement, growth, and decay attributes are valid would be beneficial for allowing a complete assessment of all CCFP attributes. This would also make the CCFP a more useful tool in delay prediction and mitigation.
2.4 Weather Trends and Analysis

Much of the literature focused on the effect of weather on air traffic. This literature does not attempt to make an analytical connection between weather and NAS performance, but instead reports the effect weather has on air traffic during specific instances of weather. Papers also focus on what products, not yet available or even developed, would have a impact on aviation efficiency.

Early work on weather and it’s effect on airport operations was done by Robinson [22]. This paper identified the weather effects which are most likely to be pertinent and the types of impacts that they have. The paper developed a methodology for isolating the weather effects on aircraft operations from other delay causing factors, indicates major delay-causing events, and provides a preliminary estimate of the impact of weather on airline operations.

The weather elements which created significant increase in delays were reduced visibility and thunderstorms. The greatest limitation of this work is that the data is only from Atlanta. Different airports in other parts of the country may have different results. This is an early work that relied on a small data set.

Weber and Evans [23] attempt to find quantitative metrics for thunderstorm operational impact. However, the research only focuses on 20 days where thunderstorms were present during the 2006 convective season. Therefore, it is really only an example of what thunderstorms can do to NAS efficiency. Metrics are developed to measure the performance of new technologies and unavoidable delay is defined as the minimum delay if all forecast information was perfect. Maximum delay reduction potential is defined as the goal for any research

Fellman and Topiwala [24] describe the influence of weather in a reroute by the ARTCC, but it does not highlight any trends on how decisions are made.

Krozel et al. [25] describes a number of classes of weather related problems that impact the efficiency of current NAS operations. It highlights significant weather effects and their primary impact on operations. The weather effects are divided into three categories;
• Surface Delays
• Terminal Delays
• En-route Delays

Within surface delays, shifting winds, low visibility, de-icing, and snow, slush, and ice on the runway have the most significant effect on operations. Within terminal delays, convective weather affects departures or arrivals, weather constraints affecting arrival/take-off tradeoffs, and weather constraints affecting arrival airspace capacity have the greatest effect. Finally within en-route delays, clear air turbulence, high top convective weather, and multiple clusters of weather cells have significant impact on operations. It then presents ideas that attempt to address the root causes. The greatest asset of this paper is to illuminate the problems associated with weather, however it does not offer any solutions.

2.5 Delay Models

The final portion of the literature search includes papers that attempt to build models to predict airline operation rates. Some describe research and methods that require a missing piece that our research provides while other attempt to predict Ground Delay Programs and delay just like our research, except through different techniques.

2.5.1 The Interactive Weather Brief

Rodenhuis [26] introduces the Interactive Weather Brief (IWB) which attempts to rationalize the estimates of the capacity with weather forecasts. The forecaster is expected to reduce the 4-dimensional weather forecast to a product for aviation users: a 1-dimensional time-sequence of capacity for selected locations/elements of the airspace (terminals, TRACONs, and routes), and to include a description of the cause of restrictions in capacity (adverse weather).

The goal of the IWB is to advise on air traffic capacity that is implied by the weather forecast, and to bring all users to the common viewpoint and appreciation of the forecast,
its skill, as well as its uncertainty. From this information traffic flow managers at the TMU will make their decisions. Our research fulfills this requirement by creating delay predictions out of forecast inputs.

### 2.5.2 Surface Conditions and Runway Configurations

Rodenhuis [11] describes the Hub forecast which utilizes terminal weather forecasts (TAF) and known surface conditions on designated runway configurations to estimate terminal capacity through Aircraft Arrival Rates (AAR). The intended users of the product are the Traffic Management Unit, the primary decision makers, and the Center Weather Service Units that are located at the Air Route Traffic Control Centers (ARTCCs).

The critical element of the Hub forecast is to connect weather variables with AARs. For this purpose, historical AARs were related to concurrent weather by first collecting the runway configurations of 8 terminal areas within the central part of the United States. Subsequently, the relationship between adverse weather and AAR is estimated from the FAA’s Operational Information System (OIS), and the results subjectively confirmed and adjusted after interviews with tower operations.

An assessment of these results and their use in practice shows that the relationship between weather and actual operating conditions (throughput) is uncertain, even when the runway selection is specified. However, the paper admits that there is plentiful data for analysis that could be used in a subsequent analysis to stratify the observations into (runway configuration, weather category, actual arrival rate). From this analysis the mean, max, and range of AAR could be determined. These improved empirical relationships would increase the value of the Hub Forecasts. Our research makes this empirical connection which would improve the research.

### 2.5.3 Regression Models

Least squares regression is a popular technique to predict future outcomes due to it’s availability in most common statistical software programs. It has been used in some of the
literature to predict delays.

Post et al. [27] developed a regression model of the overall system delay that captured the effects of both terminal and en route weather. The independent variables consisted of:

- Overall number of scheduled arrivals at the ASPM 55 airports
- The number of scheduled arrivals in IMC conditions
- The excess of scheduled arrivals over the recorded airport acceptance rates
- En route weather index that uses lightening strike data
- Flight plan data
- A number of dummy variables to capture weekend and holiday effects.

The dependent variable was the sum of the total delay. The regression resulted in a model that explained about 77% of the variation in the logarithm of the total delay.

Rehm and Klawonn [28] also applied regression which used Automated Terminal Information Service (ATIS) weather data as the independent variable and arrival time at Frankfurt Airport as the dependent variable. Arrival time meant the time it took the aircraft to land after entering the terminal area. The regression resulted in a model that explained 63% of the variation in the arrival time.

2.5.4 Bayesian Models

Pepper et al. [29] attempts to account for uncertainty in weather information at the time of TFM decisions using Bayesian Networks. The paper admits though, that the data from past TFM events in itself is not enough to distinguish between the efficiency of different strategic TFM decisions at least for delay, cancelation, diversion, and departure backlog performance metrics. Patterns in TFM performance metrics do exist, but there is wide variability across TFM events.
2.5.5 Histogram Method

Inniss and Ball [30] developed a model to take into consideration the stochastic nature of weather based on the fact that airport capacities are subject to substantial uncertainty because of weather. The main objective of the paper was to develop probabilistic capacity weather forecasts.

Given the empirical data about capacity, relative frequency histograms can be constructed and used to estimate Capacity Probabilistic Distribution Functions (CPDF). In order to account for changes over time in the underlying weather mechanism, this paper describes how they created daily, monthly or seasonal CPDF. These CDPFs are histograms with the duration of GDP in hours is binned on the $x$-axis. This generates the required input for stochastic GDP models which have the potential of improving existing decision support tools.

This study only looked at one airport (SFO). The majority of the study focused on the best possible set of clusters (monthly, seasonal, etc.) to create CPDF. The paper focus is on deriving probabilistic forecasts of arrival capacity for the Hoffman-Rifkin Model. The paper assumes that normal capacity at SFO is 45 flights per hour and that a GDP reduces that to 30 flights per hour. Therefore if you know the duration of the GDP, you will know the overall airport capacity distribution. Our research determine AARs based on historical delay data without assumed AARs determined by controllers.

2.5.6 Tree and Clustering Methods

Rehm and Klawonn [28] use Classification and Regression Trees (CART) algorithm to determine important weather factors and significant values for these weather factors which should enable them to predict travel time. Trees that are more complex can predict the training data set better and better, however complex trees yield poor results on test data. Also, our research may have better results if we use trees as a classification algorithm like the support vector machine instead of a regression like tool.

This paper compared trees to regression, but never made any conclusions as to the
advantages of one method to another. Our research hopes to illuminate this missing part.

2.6 Literature Review Summary

The literature review highlights three issues. First there are solutions to relieve delay, however they require time to implement. Second there is no research that provides a tool to produce a common operational picture. The Decision Support Tool proposed in this dissertation will provide such a tool. Finally, there needs to be a connection between the weather forecast and airport operations. This process should captures the historical stochastic properties of the AAR predictions and provide a deterministic answer the predicted delay.
Chapter 3: Methodology

The intent of this chapter is to

- introduce the Military Decision Making Process (MDMP),
- apply it to civilian air traffic management,
- create a tool to provide information needed to make the process work and
- describe the operational process to integrate the tool to improve NAS operations.

3.1 Military Decision Making Process

Successful operations for any organization, whether it be military or civilian, can only occur if all stakeholders have a common understanding of the doctrinal principles, a working standard operating procedure (SOP), and logical working thought process for examining the possibilities of the upcoming day’s operations[31]. Managers make decisions, but an organized staff can help make and communicate those decisions and ensure that they are executed. In this section we describe the decision process and the staff positions required to make it work.

3.1.1 Military Planning

Combat operations are more chaotic and contain much more uncertainty than any day in the NAS. Like air traffic management, combat is affected by the weather and operational congestion as units must be moved and monitored in restricted space. In both situations mistakes can lead to death of personnel operating within the restricted space. However, most pilots and passengers expect to survive their day of operations and do not face the
possibility of being killed by others. The effect of fear on soldiers and the actions of a human enemy are often difficult to predict. Despite the chaotic nature of combat, the military uses planning for all operations. Lack of planning is considered unprofessional and potentially deadly. The military takes planning so seriously that it is taught to all commissioned and non-commissioned officers at several times during their careers and has set up schools that teach planning at all levels.

Planning is a crucial part of the execution of any operation. Although changes occur during execution, planning provides the intent of the senior traffic flow manager so all subordinate units can anticipate what is happening next. Planning designates specific times and places as crucial to maintaining smooth traffic flow. Planning provides a framework for the collection of data inputs required for new predictive tools and organizes the outputs so they are useful to NAS users. Planning designates a list of critical information requirements that trigger future contingencies.

Planning should not be thought of a centralized inflexible exercise that takes power away from subordinate elements, but instead help these elements focus their own planning within the constraints of the daily challenges of life in the NAS. Military planning uses decision points to trigger branch plans. Decision points are a time and or place where an event may occur thatforces managers to make a decision[8]. Planning determines these decision points ahead of time and then develops a set of branch plans to deal with the possible events. This takes an infinite amount of possible courses of action and reduces it to a finite, more manageable set which can be handed down in a plan to NAS users. This provides predictability to the airlines and the traffic managers at the ARTCCs.

3.1.2 Air Traffic Management Intelligence

To create a set of decision points, ATM planners do an “enemy” estimate. What is the enemy to air traffic flow? In most cases air traffic flow is constrained due to congestion. This congestion is caused by two major components, schedule congestion and weather. Planning tools will be used to find the place and time of the congestion which will form
the set of decision points. Then branch plans will be developed to mitigate the congestion. These branch plans do not have to be new and probably should come out of a set of already agreed upon “plays”. The ATM community is already using a playbook, so this concept is not new [15].

3.1.3 Planning Components

The military uses a systematic method to plan operations with a planning staff formed to work different parts of the plan. These positions can be adapted to air traffic management. The TFM Commander (CDR), the Chief of Staff (COS), the Operations Officer (OO), the Intelligence Officer (IO), the Personnel Officer (PO), and the Resource Officer (RO) are the six principle positions of the Air Traffic Control System Command Center (ATCSCC) Staff and is depicted in Figure 3.1. The CDR and the COS perform supervisory tasks over the entire ATCSCC Staff while the four other positions supervise subordinate staffs that collect information, run models, and provide critical output products.
Traffic Flow Management Commander

The TFM Commander is in charge of the planning process. From start to finish the commander’s personnel role is central to ensuring that the staff meet the requirements of time, planning horizons, simplicity and level of detail. The CDR does this by visualizing, describing, and directing operations [8]. The CDR should be an experienced air traffic manager with central authority to approve all decisions that affect the NAS.

Chief of Staff

The staff’s effort during planning focuses on developing an effective plan. It does this by integrating information with technical competence. The Chief of Staff manages, coordinates, and focuses the staff’s work, and provides quality control. The COS must clearly understand the commander’s guidance and intent because they supervise the entire process. The COS provides time lines to the staff, establishes backbrief times and locations, and provides any instructions necessary to complete the plan. The COS must be an experienced traffic flow manager who is the second in command and therefore must have the authority to act in absence of the CDR.

Operations Officer

The Operations Officer is in charge of the execution of the plan. The OO must be aware of all scheduled airline traffic, Air Route Traffic Control Center (ARTCC), and airport operations and issues. This massive task is not done alone. The OO should have a large enough staff to keep up with all required information. The OO must also be able to rely on the airlines and the ARTCCs for updated information. The OO’s principle focus should be on en route and terminal area congestion and the actions proposed to mitigate any delay with an acceptable level of risk. The OO also should have models at his disposal to predict congestion and an available playbook with mitigation actions. The OO is the third in command who must have the authority to approve all routine NAS actions in the absence of the CDR or the COS.
**Intelligence Officer**

The Intelligence Officer is in charge of maintaining awareness of all possible obstacles to smooth traffic flow. The IO should work closely with the OO to identify congested areas. With the increase in technology, one of the most important jobs of the IO will be to collect required input information, run predictive models, and organize the output in a way so it is useful to the CDR to make decisions. The principle cause of delays is weather, so the largest part of the IO’s work should be focused on weather prediction and the impact on the NAS. The IO should help the CDR identify critical points of location or time that could affect traffic flow and continue to monitor these points during execution. Initial planning information is important, but continuing updates are also critical in executing planned operations.

**Personnel Officer**

The Personnel Officer monitors the staffing at the ARTCC’s. The PO advises the ARTCCs on potential congestion that may increase the need for more controllers to handle the increased workload. The PO also informs the staff whether or not the staffing at particular ARTCCs may decrease the traffic through the sector. The POs knowledge of ARTCC staffing is crucial to ensuring that plans formulated can actually be executed.

**Resource Officer**

The Resource Officer monitors the equipment availability of the airlines and ATC equipment maintenance issues. Planes grounded due to maintenance may change departure times which could have an effect on congestion predictions. Also ATC equipment may be down due to maintenance or upgrade, which could lead to reduced capacity within the affected sectors. The job of the RO is to maintain contact with the airline and ATC maintenance and work together with other members of the staff to determine the effect of a mechanical break down.
3.1.4 Planning Tools

Except the commander and the chief of staff, each member of the planning staff is in charge of a section staffed by enough personnel to complete the required planning tasks. Within these sections, tools are required to help collect and analyze data. The Resource Section and the Personnel Section will have to maintain a updated database filled with the required personnel and resource requirements. For instance the Personnel Section’s database should have the up to date duty rosters for all Air Route Traffic Control Centers (ARTCC) while the Resource Section should have up to the minute information on aircraft maintenance. The biggest barrier to both these databases would be the release of sensitive information from the airlines. The software though is available and can be purchased off the self.

As discussed earlier, the Operations Section should have a tool to model the current airline schedule and predict possible congestion points. At present there are tools available such as the Sector Design Analysis Tool (SDAT). SDAT is intended to provide the airspace designer with a fast, easy, and accurate way to develop and evaluate proposed changes to airspace structure and/or traffic loading [32]. Predictions of conflict potential, traffic and airspace loading, and impacts on the airspace user can be generated for any proposed combination of airspace and traffic data. These predictions highlight potential congestion delays which allow the planners to propose branch plans to counter these conflicts. Scheduled congestion can also be predicted by using simulation tools that fly the current commercial airline schedules such as the Future ATM Concepts Evaluation Tool (FACET) developed by NASA [33].

The last remaining section is the Intelligence Section. In the military, the Intelligence Section is in charge of enemy actions or any condition that would impede friendly progress. What is the enemy to air traffic management? The answer is congestion which is caused by two major issues; schedule congestion and poor weather conditions. Schedule congestion can be highlighted using available tools such as SDAT and FACET describe above. At present though, there are few weather tools available to provide predictions of the time and the magnitude of delays.
3.1.5 Decision Point Development

Facts and assumptions are collected and processed by the planning staff. The planning staff is crucial in producing the plan for the NAS day. Each member of the staff should be an expert in the area and well versed in the entire process so they understand their individual role. The Chief of Staff runs the planning process and ensures it is done correctly by maintaining staff discipline. After identifying the facts available, the staff must develop assumptions. Assumptions replace necessary but missing or unknown facts. Validity defines whether or not an assumption is likely to occur. “Assuming away” potential problems is the example of an invalid assumption[31].

Each of the section officers prepare estimates to be brief to the ATC commander. The ATC commander is responsible for all national ATC operations. In the planning process, the commander is responsible for focusing the efforts of the staff on essential information that is needed to make informed, timely, but not necessarily perfect decisions. The staff estimates should focus on critical information that could affect operations. This information is the genesis of the day’s decision points. A decision point is an event, area, or point in the NAS where the commander must make a critical decision. A decision point requires a decision by the commander. It does not dictate what the decision is, only that the commander must make one, and when and where it should be made to have the maximum impact on NAS operations [8]. For instance, if the weather brief indicates that predicted weather over EWR may cause delays at 2100 Zulu, then planning must set a time that the commander must decide to execute one of the proposed branch plans to deal with predicted irregular operations.

The result of effective planning should be a collection of decision points. With each decision point should be a set of branch plans that offer the commander a finite set of solutions to obstructed flow. Good planning should highlight the consequences of all the actions, so the commander can make the most well informed decision. Since planning is not a rigid process, the commander may choose one plan, a combination of several plans, or decide to do nothing.
After the decision points and associated branch plans have been developed, revised, and approved they should be published and sent to subordinate organizations and the airlines. Now all stakeholders within the NAS have a predictable set of possible actions that may be executed by the command center. They also can now do their own planning based on how these plans will affect their individual operations.

3.2 Weather Delay Prediction Tool

As stated earlier in the section on military planning, at present there is not tool to predict future NAS performance based on predicted weather. For efficient operation of the NAS, there is a need for the weather forecasting services and TFM products to estimate the reduction in capacity due to adverse weather. Weather forecast products are uncertain and the uncertainty increases with lead-time. Useful applications of weather forecasts requires either refinement, consultation, and application of the weather forecast to estimate air traffic capacity or decision support tools that take forecasts and make predictions based on past forecasts and those forecasts connections to NAS capacity [11]. This paper describes a methodology used to create one such decision support tool known as the Weather Delay Prediction Tool. With this tool, the user enters the TAF for a given day and airport and the tool provides AAR predictions which can be derived to estimate delay and Ground Delay Program (GDP) time and duration.

Initially, this research focused on the Collaborative Convective Forecast Product (CCFP) as the weather forecast. The CCFP is a thunderstorm forecast for the entire United States and Canada and the research focused on its use as a predictive tool. After conversations with traffic management personnel and airline management, it was concluded that they rarely used the CCFP for any weather planning and relied on the Terminal Aerodrome Forecast (TAF) instead. The TAF has a good collection of available archived forecasts, so it was a good fit for the research objectives. To measure delays, a tool to predict GDPs was first considered. Over the course of the research it was determined that measuring delays may be more appropriate and then derive GDPs from the results. However, after presenting
the work to some air traffic management experts, it was determined that it was better to use Aircraft Arrival Rates (AAR), since that was a common used factor to measure degraded airport capacity due to irregular operations. Also, GDPs and delays can be derived easily if the AAR is known. The general procedure used to determine a connection was:

- Collect data from the various available data sources,
- using assorted tools, format the data into a usable layout,
- use a classification tool to connect the two sets, and
- test the data to ensure there is a correlation.

### 3.3 Data Sources

To test this theory, the following 8 airports were selected to collect data:

- Newark Liberty International (EWR)
- Chicago O’Hare International (ORD)
- Atlanta Hartsfield International Airport (ATL)
- Philadelphia International Airport (PHL)
- Reagan National Airport (DCA)
- Dulles International Airport (IAD)
- LaGuardia Airport (LGA)
- John F. Kennedy International Airport (JFK)

These airports were selected because they are either major hubs or they are inside the busy northeast corridor. These northeast airports are not only economically important, but also politically important due to their location near high population areas. The data used in this paper came from three areas:
• The TAF data was collected from a website provided by the National Climatic Data Center (NCDC).

• The Aircraft Arrival Rate data was collected from the Aviation System Performance Metrics (ASPM) database based maintained by the FAA.

• The delay data was found on the Bureau of Transportation Statistics website for summary statistics for destination airports.

3.3.1 Ground Delay Program

One of the features of the weather delay prediction tool is to predict a Ground Delay Program (GDP) and its duration. GDPs are implemented to control air traffic volume to airports where the projected traffic demand is expected to exceed the airport’s arrival rate for a lengthy period of time. Lengthy periods of demand exceeding acceptance rate are normally a result of the airport’s arrival rate being reduced for some reason. The most common reason for a reduction in arrival rate is adverse weather such as low ceilings and visibility.

Flights that are destined to the affected airport are issued Controlled Departure Times (CDT) at their point of departure. Flights that have been issued CDTs are not permitted to depart until their Controlled Departure Time. These CDTs are calculated in such a way as to meter the rate that traffic arrives at the affected airport; ensuring that demand is equal to arrival rate. The length of delays that result from the implementation of a Ground Delay Program is a factor of two things; how much greater than the arrival rate the original demand was, and for what length of time the original demand was expected to exceed the arrival rate [34].

3.3.2 Terminal Aerodrome Forecast

The TAF is an operational forecast consisting of the expected meteorological conditions significant to a given airport or terminal. TAFs always include a forecast of surface wind
speed and direction, visibility, and clouds. Weather type, obstructions to vision, and low level wind shear are included as needed. The National Weather Service (NWS) produces over 570 TAFs.

NWS is responsible for providing terminal forecasts to commercial and general aviation pilots for the protection of life and property and in response to requirements levied by International Civil Aviation Organization (ICAO) via the FAA in order to promote the safety and efficiency of the National Airspace System (NAS).

**Forecast Production**

A TAF is a report established for the 5 statute mile radius around an airport. In the U.S., TAFs are produced four times a day starting at approximately 30 minutes before each main synoptic hour (00Z, 06Z, 12Z, and 18Z). All the forecasts produced starting one hour before the main synoptic hour up to four hours past the main synoptic hour are considered to be for the same cycle. For example, forecasts produced between 1100Z and 1600Z are all considered to be 12Z forecasts. In reality, forecasts contain the starting and ending times for which the forecast is valid. Our rule of thumb is that all forecasts produced between (hh-1)00Z to (hh+4)44Z are considered to be for the hh cycle. Table 3.1 summarizes the relationship between forecast production time and cycle [35].

<table>
<thead>
<tr>
<th>Forecast Cycle</th>
<th>Forecast Reduction Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>00Z</td>
<td>2300 - 0459</td>
</tr>
<tr>
<td>06Z</td>
<td>0500 - 1059</td>
</tr>
<tr>
<td>12Z</td>
<td>1100 - 1659</td>
</tr>
<tr>
<td>18Z</td>
<td>1700 - 2259</td>
</tr>
</tbody>
</table>
TAF Format

The TAF is reported in a standard international format code and must follow a set of rules that define what must be placed in each line and what criterion requires a new line. A TAF report contains the following sequence of elements in the following order:

- Type of Report
- ICAO Station Identifier
- Date and Time of Origin
- Valid Period Date and Time
- Forecast Meteorological Conditions

The report type header will always appear as the first element in the TAF forecast. There are two types of TAF reports, a routine forecast, TAF, and an amended forecast, TAF AMD. An amended TAF is issued when the current TAF no longer adequately describes the ongoing weather or the forecaster feels the TAF is not representative of the current or expected weather.

The TAF code uses the ICAO four-letter location identifiers. In the conterminous United States, the three-letter identifier is prefixed with a K. For example SEA (Seattle) becomes KSEA. Elsewhere, the first one or two letters of the ICAO identifier indicate in which region of the world and country (or state) the station is. Pacific locations such as Alaska, Hawaii, and the Marianas islands start with P followed by an A, H, or G respectively. The last two letters reflect the specific station identification. If the location’s three-letter identification begins with an A, H, or G, the P is just added to the beginning. If the location’s three-letter identification does not begin with an A, H, or G, the last letter is dropped and the P is added to the beginning.

The next element is the UTC date and time the forecast is actually prepared. The format is a two-digit date and four-digit time followed, without a space, by the letter Z. Routine
TAFs are prepared and filed approximately one-half hour prior to scheduled issuance times. TAFs are scheduled for issuance four times daily at 0000Z, 0600Z, 1200Z, and 1800Z.

The UTC valid period of the forecast is a two-digit date followed by the two-digit beginning hour and two-digit ending hour. Routine TAFs are valid for 24-hours. Valid periods beginning at 0000Z shall be indicated as 00. Valid periods ending at 0000Z shall be indicated as 24. The 24 indication applies to all time group ending times. In the case of an amended forecast, or a forecast which is corrected or delayed, the valid period may be for less than 24 hours.

This forecast meteorological conditions is the body of the TAF. The basic format is:

- Wind
- Visibility
- Weather
- Sky Condition
- Optional Data

The wind, visibility, and sky condition elements are always included in the initial time group of the forecast. Weather is included in the initial time group only if significant to aviation. If a significant, lasting change in any of the elements is expected during the valid period, a new time period with changes is included. The new time period will include only those elements which are expected to change; i.e., if a lowering of the visibility is expected but the wind is expected to remain the same, the new time period reflecting the lower visibility would not include a forecast wind. The forecast wind would remain the same as in the previous time period.

The wind group includes forecast surface winds. The surface wind is the expected wind direction (first three digits) and speed (last two or three digits if 100 knots or greater). The contraction KT follows to denote the units of wind speed in knots. Wind gusts are noted
Table 3.2: Weather Forecast Codes

<table>
<thead>
<tr>
<th>Intensity or Proximity</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>MI</td>
</tr>
<tr>
<td>+</td>
<td>BC</td>
</tr>
<tr>
<td>VC</td>
<td>DR</td>
</tr>
<tr>
<td></td>
<td>BL</td>
</tr>
<tr>
<td></td>
<td>SH</td>
</tr>
<tr>
<td></td>
<td>TS</td>
</tr>
<tr>
<td></td>
<td>FZ</td>
</tr>
<tr>
<td></td>
<td>PR</td>
</tr>
<tr>
<td></td>
<td>BR</td>
</tr>
<tr>
<td></td>
<td>FG</td>
</tr>
<tr>
<td></td>
<td>FU</td>
</tr>
<tr>
<td></td>
<td>DU</td>
</tr>
<tr>
<td></td>
<td>SA</td>
</tr>
<tr>
<td></td>
<td>HZ</td>
</tr>
<tr>
<td></td>
<td>PY</td>
</tr>
</tbody>
</table>

**Intensity or Proximity**

- by the letter G appended to the wind speed followed by the highest expected gust (two or three digits if 100 knots or greater).

**Expected prevailing visibility** is forecast in statute miles and fractions of statute miles followed by SM to note the units of measure. Statute miles followed by fractions of statute miles are separated with a space, for example, 1 1/2SM. Forecast visibility greater than 6 statute miles is indicated by coding P6SM. Directional or variable visibility is not forecasted and the visibility group is omitted if missing.

The expected weather phenomenon or phenomena is coded in TAF reports using the format in Table 3.2.

If no significant weather is expected to occur during a specific time period in the forecast, the weather group is omitted for that time period. If, after a time period in which significant weather has been forecast, a change to a forecast of no significant weather occurs, the
contraction NSW (No Significant Weather) will appear as the weather group in the new time period. However, NSW is only included in the BECMG or TEMPO groups.

The FM group is used when a rapid change, usually occurring in less than one hour, in prevailing conditions is expected. Typically, a rapid change of prevailing conditions to more or less a completely new set of prevailing conditions is associated with a synoptic feature passing through the terminal area (cold or warm frontal passage). Appended to the FM indicator is the four-digit hour and minute the change is expected to begin and continues until the next change group or until the end of the current forecast.

A FM group will mark the beginning of a new line in a TAF report. Each FM group contains all the required elements – wind, visibility, weather, and sky condition. Weather will be omitted in FM groups when it is not significant to aviation. FM groups will not include the contraction NSW.

The BECMG group is used when a gradual change in conditions is expected over a longer time period, usually two hours. The time period when the change is expected is a four-digit group with the beginning hour and ending hour of the change period which follows the BECMG indicator. The gradual change will occur at an unspecified time within this time period. Only the conditions are carried over from the previous time group.

The TEMPO group is used for any conditions in wind, visibility, weather, or sky condition which are expected to last for generally less than an hour at a time (occasional), and are expected to occur during less than half the time period. The TEMPO indicator is followed by a four-digit group giving the beginning hour and ending hour of the time period during which the temporary conditions are expected. Only the changing forecast meteorological conditions are included in TEMPO groups. The omitted conditions are carried over from the previous time group [35].

**TAF Example**

This TAF example in Figure 3.2 is from Calgary International Airport, Calgary, Alberta, and was released on October 19, 2006 at 2038 UTC:
TAF CYYC 192038Z 192118 17008KT P6SM SCT020 OVC080 TEMPO 2203 P6SM -SHRA
BECMG 2223 24007KT
FM0300Z 32010KT P6SM SCT007 BKN060
FM0600Z 33015KT P6SM SCT010 BKN040 TEMPO 0812 5SM -RASN BR OVC010
FM1200Z 34015G25KT P6SM SCT010 OVC030 TEMPO 1218 2SM -SHSN OVC010

Figure 3.2: Example TAF

- TAF indicates that the following is a terminal area forecast.
- CYYC indicates that the report came from Calgary International Airport.
- 192038Z indicates that the report was issued at 2038 UTC on the 19th of the month.
- 192118 indicates that the report is valid from 2100 UTC on the 19th until 1800 UTC on the following day.
- 17008KT indicates that the wind is forecasted in the first part of the forecast (2100 to 0300 UTC) to be from 170 degrees at 8 knots.
- P6SM indicates that visibility is forecasted to be at least six statute miles. Forecasted visibility of six miles or more is always referred to as P6SM.
- SCT020 OVC080 indicates that clouds are forecasted to be scattered at 2000 feet and overcast at 8000 feet.
- TEMPO 2203 P6SM -SHRA indicates that between 2200 and 0300 there may be at times light rain showers with visibility of at least six statute miles.
- BECMG 2223 24007KT indicates that a wind shift to 240 degrees at 7 knots is forecasted to occur between 2200 and 2300 UTC.
- FM0300Z 32010KT P6SM SCT007 BKN060 indicates that beginning at 0300 UTC the wind will be from 320 degrees at 10 knots, visibility will be at least six statute miles, and clouds will be scattered at 700 feet and broken at 6000 feet.
• FM0600Z 33015KT P6SM SCT010 BKN040 TEMPO 0612 5SM -RASN BR OVC010 indicates that beginning at 0600 UTC the wind will be from 330 degrees at 15 knots, visibility will be at least six statute miles, and clouds will be scattered at 1000 feet and broken at 4000 feet. There is forecasted to be at times between 0600 and 1200 hours visibility at 5 statute miles, rain showers, snow showers, and mist with an overcast layer of cloud at 1000 feet.

• FM1200Z 34015G25KT P6SM SCT010 OVC030 TEMPO 1218 2SM -SHSN OVC010 indicates that beginning at 1200 UTC the wind will be from 340 degrees at 15 knots gusting to 25 knots, visibility will be at least six statute miles, and clouds will be scattered at 1000 feet and overcast at 3000 feet. There is also forecasted to be at times between 1200 and 1800 hours visibility of two statute miles, light snow showers, and an overcast layer of cloud at 1000 feet [36].

3.3.3 Aircraft Arrival Rates

A Strategic Plan of Operations for managing flows during severe weather events in the National Airspace System (NAS) takes into account reduced Airport Arrival Rates (AARs) due to weather constraints. If the predicted capacity (number of aircraft that the airport can safely land in a given time period) falls short of scheduled demand (number of aircraft that wish to land at an airport in a given time period), traffic flow managers may implement a Ground Delay Program (GDP) [37]. GDPs are implemented by the Air Traffic Control System Command Center (ATCSCC) after consultation with regional Federal Aviation Administration (FAA) centers and with airline operations centers. A GDP applies to a particular airport, has specified start and stop times, and sets an allowable arrival rate.

Originally this research we focused on predicting GDPs by using the SVM. However, after discussions with air traffic managers, it was decided that it was more appropriate to predict AARs. AARs offer several advantages. First, each airport tends to revert to a finite set of AAR rates when airport capacity had to be reduced due to weather. This allowed grouping the possible outcomes into only a few distinct bins. Then a value was
chosen between each bin and tested whether the day was ≥ to the in between value or < the between value. Finally, a predictor function was developed for each of these values and from the results we were able to predict the future AAR.

The second advantage of the AAR was that GDPs could be predicted based on the conclusions of the predictor function. GDPs occur when the AAR is below the rate for a normal operations when the weather is favorable. AAR predictions are made for four times during the day based on the demand level of the airport. This generated a graph found in Figure 3.3. For EWR, the greatest demand hours were at 0700, 1100, 1500, and 2000 local time. Table 3.3 shows the demand hour and the assumed coverage hours for EWR. EWR’s normal AAR was 44, so Figure 3.3 predicts a GDP from 1300 to 2400.

The ASPM database provides the hourly AAR rates and demand for all major airports. To determine the best time to collect AARs, the busiest time periods at an airport had
to be determined. The average was found using the ASPM data for each hour of the day using data from January 2002 through June 2007 [38]. Figure 3.4 shows the average hourly demand for Dulles International Airport (IAD). Four peaks are observed in the graph which indicate the busiest time of day at the airport, which in this case is at 0700, 1100, 1500, and 2000 local time. Therefore, from the collected ASPM data, AARs were collected for those specific time periods of the day to serve as our dependent variables for the analysis. A binary classifier is used as part of our prediction algorithm, so a set of common AARs must be chosen to create predictor functions. Fortunately, the tower controllers at most of the major airports tend to fall back on only a small set of AARs. For instance at IAD, the most common AAR is 90 per hour. If conditions start to deteriorate, the IAD tower most commonly reduces the AAR to either 80, 75, or 60.

3.3.4 Delay Data

The final data collected was the delay statistics for individual airports. We found this data on the Bureau of Transportation Statistics website for summary statistics for destination airports. From the website we collected Total Number of Flights, Arrival Delay (minutes),
and Arrival Time for each day. With this information, we then found the average daily delay and the delay for the time periods 0600 to 1000, 1000 to 1400, 1400 to 1800, and 1800 to 2200 local time [39].

3.4 Support Vector Machines

The Support Vector Machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. In our case we are using the mapping function as a classification function. In addition to its solid mathematical foundation in statistical learning theory, SVMs have demonstrated highly competitive performance in numerous real-world applications, such as bioinformatics, text mining, face recognition, and image processing [40]. SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the Figure 3.5.

In this example, the objects belong either to class square or circle. The separating line defines a boundary on the right side of which all objects are squares and to the left of which all objects are circles.

Figure 3.5 is a classic example of a linear classifier, i.e., a classifier that separates a set
of objects into their respective groups (square and circle in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (training cases). This situation is depicted in Figure 3.6.

Compared to Figure 3.5, it is clear that a full separation of the square and circle objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper-plane classifiers. Support Vector Machines are particularly suited to handle such tasks.

Figure 3.7 shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all one has to do is to find an optimal line that can separate the square and the circle objects.
Support Vector Machine (SVM) is a method that performs classification tasks by constructing hyper-planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either -1 or 1. For this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

s.t.  
$$y_i (w^T x + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1 \ldots N$$

Where $C$ is the capacity constant, $x$ is the vector of coefficients, $b$ a constant, $y$ the dummy variable, and $\xi_i$ are parameters for handling non-separable data (inputs). The index $i$ labels the $N$ training cases. The kernel $w$ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the $C$, the more the error is penalized. Thus, $C$ should be chosen with care to avoid over fitting [41].
3.5 Proposed Method

After collecting the TAF data as the independent variable matrix and the ASPM AAR as the dependent variable vector the SVM was applied to determine a function to predict future AAR’s. The quadratic program introduced earlier was coded into AMPL. AMPL is a comprehensive and powerful algebraic modeling language for linear and nonlinear optimization problems, in discrete or continuous variables. After coding, the program was submitted and the associated data to the NEOS Server for Optimization. The AMPL code is found in Appendix A.

3.5.1 Creating the TAF Vector

In this specific case, the $x$ in the quadratic program represented the 57 character long vector from the TAF weather data collected from 2002 through 2006. To create the vector, TAF data was collected from a website provided by the National Climatic Data Center (NCDC). Appendix E shows a website screenshot of the HDSS Access System. The HDSS is the robotic tape assembly used to store large datasets at NCDC. The data must be transferred from the tapes onto the public ftp site. The HAS web interface allows users to order TAF data from the tape archive. These files tend to be long, up to 100 pages of text data, because all reports received are placed in these files as they are received and they are updated approximately every five minutes as data becomes available. Also, forecasts may be duplicated within the files and multiple forecasts received from a station may appear in a file [42].

To transform the raw TAF data into usable vector form, data was pasted into an Excel Spreadsheet. Then the text to column function was used to put each part of the data into a separate cell. Then columns were created to find times or selected weather codes from the raw data. There are two important times for this research. The first is the time the forecast was created and the second is the time the forecasted weather is to occur. In this spreadsheet, the first column contained a function to find the date of the specific piece of data. The next column found the time of day and the third column combined the data
together to form a date time group value. The fourth column searched for the time which the forecasted weather was to occur.

After the report and forecast times were determined the next task was to find the specific weather forecast parameters. One column each was used to find the wind direction, wind speed, and visibility for the forecast period. The next two columns checked for codes that indicate rain and snow, while the next five columns further described the precipitation by searching for codes that indicate showers, thunderstorms, fog, mist, or freezing precipitation. The final four columns extracted ceiling data.

After the data was transformed into a linear format, it was then parsed down to include only the 0600 Zulu TAF reports. It was assumed that planning would take place early in the morning and the 0600 Zulu TAF, which equates to 0100 EST, was the first of the day. Then the wind direction was examined to determined if it was a cross wind. Whether or not there was a cross wind was determined by first obtaining the airport map as seen in Figure 3.8.

Figure 3.8 shows that the primary runways at Philadelphia are 9L/R and 27 L/R. This
means that the winds commonly flow in an east-west direction. Therefore we assumed a cross wind if the wind is blowing from a direction of 320 degrees to 40 degrees or 140 degrees to 220 degrees. So, if from the extracted data, a wind direction was from those directions, a binary value of 1 was entered. The final step is to find the 0600 Zulu TAF forecasts for 1100 Zulu, 1500 Zulu, 1900 Zulu, and 2300 Zulu. This creates a vector with 56 values. Because there has been an increase in the average delay every year, an extra 57th value was added with the last digit of the year.

3.5.2 Method

The first step in the process was to find the common AARs for each airport in the study. Using the ASPM database, AARs were collected for each of the four peak hours for the 1826 days in the dataset. Airports tend to have a set of common AARs that they use. For instance IAD maintains a rate of 90 22% of the time, 80 19% of the time, and 75 20% of the time. This provides consistent points to perform the classification algorithm.

For this case the $y$ represents a binary variable that indicates whether or not an AAR was set at a certain numerical rate for a given airport. Values equal to -1 indicate that day was greater than or equal to the numerical rate while values equal to 1 indicate that day was less than the numerical rate. One advantage of the SVM method is the way it deals with data outliers. For most methods, statistical techniques are used to eliminate values that are considered abnormalities. The SVM has an error function in the objective function, were the $C$ variable is set to a value that increases or decreases the number of incorrect classifications within the data. A high $C$ allows fewer outliers, while a smaller $C$ allows more. For our analysis $C$ was set at 1000 after experimenting with other values. This helped to determine a $\xi$ vector, which was only used to relax the function, so a solution was possible. The $\xi$ vector was not used in the final prediction function.

For the independent variables, the five years worth of data included 1826 days so this created an 1826 by 57 data matrix for the independent variable. The AMPL code was run on the NEOS Server and found a solution vector $w$ and variable $b$ for each airport. After
determining the $w$ vector and the $y$ variable the current TAF forecast could be used to develop an $x$ vector using that data and then use the Equation 3.1 to develop a prediction value.

$$w^T x_i + b$$

(3.1)

If the prediction value was greater than 0, then the algorithm predicts that less than an AAR will occur on that day. Conversely, if the value is less than zero then the algorithm predicts greater than an AAR value on that day.

### 3.6 Alternative Methods

The linear classifier presented in the earlier section has it’s limits, since it assumes that the data is separated by a linear hyperplane when some separation functions are nonlinear hypersurfaces. Fortunately, the method presented above can be extended to create nonlinear decision boundaries. The motivation for such an extension is that an SVM that can create a nonlinear decision hyperspace will be able to classify nonlinearly separate data. This will be achieved by considering a linear classification in the so-called feature space [40].

#### 3.6.1 Duality

Duality plays an important role in classification problems using support vector machines. To show this we first begin with a quadratic programming problem that assumes our data is not separable. This problem described in Equation 3.2 through Equation 3.4 determines an $n$-dimensional vector $w$ and a scalar $b$. 
Here $C$ is a penalty coefficient representing the trade off between misclassification and a wide separation margin. To define the dual of the problem it will be convenient to represent the problem constraints in matrix-vector form. Denote by $X$ the $n \times k$ matrix whose columns are the training vectors $x_i$. Let $Y$ be the diagonal matrix whose $i$th diagonal term is $y_i$ and let $e = (1, \ldots, 1)^T$ be a vector of length $k$ whose entries are all equal to one. The new formulation is shown in Equation 3.5 through Equation 3.7

\begin{equation}
\min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{k} \xi_i \tag{3.2}
\end{equation}

\begin{equation}
\text{s.t.} \quad y_i (w^T x + b) \geq 1 - \xi_i \quad \tag{3.3}
\end{equation}

\begin{equation}
\xi_i \geq 0, i = 1 \ldots N \quad \tag{3.4}
\end{equation}

Let $\alpha$ and $\eta$ be the $k$-dimensional vectors of Lagrange multipliers corresponding to the hyperplane constraints and the constraints $\xi_i \geq 0$ respectively.

\begin{equation}
\max \quad \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \alpha^T Y X^T X Y \alpha \tag{3.8}
\end{equation}

\begin{equation}
\text{s.t.} \quad \sum_{i=1}^{k} y_i \alpha_i = 0 \quad \tag{3.9}
\end{equation}

\begin{equation}
\alpha_i + \eta_i = C \quad \tag{3.10}
\end{equation}

\begin{equation}
\alpha, \eta \geq 0 \quad \tag{3.11}
\end{equation}
The problem can be written just in terms of $\alpha$.

\[
\begin{align*}
\max & \quad \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \alpha^T Y X^T Y \alpha \\
\text{s.t.} & \quad \sum_{i=1}^{k} y_i \alpha_i = 0 \\
& \quad 0 \leq \alpha_i \leq C
\end{align*}
\]

(3.12)

(3.13)

(3.14)

The dual formulation found in Equation 3.12 through Equation 3.14 offers several benefits. Firstly the dual, like the primal, is a quadratic problem, but it is easier to solve since the formulation has only equality with all other constraints being simple upper and lower bounds. The second benefit is that a binding hyperplane constraint corresponds to any positive $\alpha_i$. This allows us to determine the set of support vectors. The greatest advantage of the dual formation though, is it allows us to expand the power of the support vector machines to data that is not linear separable [44].

### 3.6.2 Nonlinear Hyperplane

Now that the dual formulation has been determined, so now a non-linear hyperplane can be fit instead of a linear one. If each two-dimensional data vector $x$ is transformed to the higher dimensional “feature” vector $\Phi(x)$, then a three dimensional hyperplane of the form $\Phi(x)^T w + b$ that separates the feature vectors corresponding to the data can be considered. This idea could be determined to define other separating nonlinear function, such as polynomials.

One issue with higher order polynomials is that the problem becomes larger due to the dimensionality. Fortunately, the dual formulation offers an approach for efficient computation. Suppose the lower dimensional points $x$ is mapped to the higher dimensional points $\Phi(x)$ and then use a separating hyperplane for the $\Phi(x)$ points in feature space. The dual problem introduce above is where $\Phi(X)$ is the matrix whose columns are $\Phi(x_i)$. 

48
Table 3.4: Possible Admissible Kernels

<table>
<thead>
<tr>
<th>Kernel Function</th>
<th>Type of Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa(x, x_i) = (x^T x_i)$</td>
<td>Linear, dot product, kernel</td>
</tr>
<tr>
<td>$\kappa(x, x_i) = [(x^T x_i) + 1]^d$</td>
<td>Complete polynomial of degree $d$</td>
</tr>
<tr>
<td>$\kappa(x, x_i) = e^{\frac{1}{2}[-x_i]^T \sum^{-1}(x-x_i)]}$</td>
<td>Gaussian</td>
</tr>
</tbody>
</table>

The only way the data appears in the dual problem is through the matrix $\Phi(X)^T \Phi(X)$ whose elements are $\kappa(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$. A kernel of the machine provides an efficient way to compute inner products of the form $\kappa(x, z) = \Phi(x)^T \Phi(z)$ for any two points $x$ and $z$ without going through the trouble of forming $\Phi(x)$ and $\Phi(z)$. This method is interested in kernel functions that can be computed efficiently without constructing $\Phi$. In fact the specific form of $\Phi$ is not a concern. Kernels must be functions where the matrix of all inner products of any number of points in the data space must be positive semi-definite [40]. Table 3.4 lists some of the possible kernel functions.

### 3.6.3 Modified Method

Using the dual formulation allows expansion to nonlinear hyperplanes, however the procedure required to produce classification predictions is not as simple. The first step is to compute inner products of the form

$$\kappa(x, z) = \Phi(x)^T \Phi(z)$$ (3.15)

for any two points $x$ and $z$. This was done using the MATLAB code found in Appendix B. Five years of training data led to 1826 distinct training days, which led to a $1826 \times 1826$ inner products matrix. The next step was to create the matrix $Y \kappa(x, z) Y$ which was programmed into AMPL code and run in the NEOS server.

Due to the large size of the matrix, the size of the data file for the AMPL code exceeded 39,000 KB, which was huge in comparison to the primal formulation which only had a data
file of around 240 KB. The large data file size became a problem because it exceeded the capacity of the NEOS server and was rejected. To avoid this a smaller file was created with a smaller data set. Of the 1826 possible days, 500 were randomly chosen to be formed into a matrix which reduced the size of the data file to 3000 KB which was able to be loaded into NEOS.

The quadratic programming problem from the dual formulation produced a vector of $\alpha$. See Appendix C for the AMPL code. From complementary slackness, any $\alpha_i$, that is positive corresponds to a binding hyperplane constraint, hence the corresponding point $x_i$ is a support vector. The set of support vectors is denoted as $SV$. The coefficients of the hyperplane $w$ can then be computed as

$$w = \sum_{i \in SV} \alpha_i y_i x_i.$$  \hfill (3.16)

Complementary slackness also implies that if $\alpha_j < C$ then $\xi_j = 0$. Thus any point $x_i$ for which $0 < \alpha_j < 0$ satisfies $y_j (w^T x_j - b) = 1$. Any such point can be used to compute the value of $b$:

$$b = w^T x_j - y_j = \sum_{i \in SV} \alpha_i y_i x_i^T x_j - y_j.$$  \hfill (3.17)

If there are several such points the average computed value of $b$ is commonly taken to ensure the highest accuracy [44]. Now that $\alpha$ and $b$ have been solved, the training data and the testing data was then classified using

$$w^T \Phi(x) + b = \sum_{i \in SV} \alpha_i y_i \kappa(x_i, x) + b.$$  \hfill (3.18)
To create a prediction tool only the identified support vectors and the value for $b$ were included in Equation 3.18. Just like the linear method, if the result is less than zero, then the classification is considered in the $-1$ class and if the result is greater than zero it is classified in the $+1$ class.

### 3.7 Performance Measures

To measure the performance of the SVM, the concepts sensitivity and specificity are often used; these concepts are readily usable for the evaluation of any binary classifier. For some time periods in the 5 year data set the SVM indicated a certain AAR, and the SVM confirms this. These instances are called true positives. Some time periods have a given AAR, but the SVM indicates that another AAR. They are called false negatives. Some SVMs indicate that an specific AAR does not occur, and the SVM confirms this - true negatives. Finally, there may not be a specific AAR indicated yet the has a positive SVM result - false positives. Thus, the number of true positives, false negatives, true negatives, and false positives add up to 100% of the set.

#### 3.7.1 Evaluation of Binary Classifiers

Sensitivity is the proportion of days that tested positive of all the positive days tested as seen in Equation 3.19.

$$Sensitivity = \frac{truepositives}{truepositives + falsenegatives}$$  \hfill (3.19)

It can be seen as the probability that the SVM indicates an AAR given that there was that specific AAR predicted. The higher the sensitivity, the fewer real AAR values go undetected. Specificity is the proportion of days that tested negative of all the negative days tested as seen in Equation 3.20.
Specificity = \frac{\text{truenegatives}}{\text{truenegatives} + \text{falsepositives}}

As with sensitivity, it can be looked at as the probability that the SVM indicates a negative for a given AAR given that AAR did not occur. The higher the specificity, the fewer negative results are labeled as positive.

In theory, sensitivity and specificity are independent in the sense that it is possible to achieve 100%. In practice, there often is a trade-off, and you can't achieve both. This is because much of the characteristics identified to determine whether a sample gives a positive or negative test may not be as obvious. If high sensitivity is desired, a very low threshold could be set which would consequently declare many GDPs. Therefore, the number of true positives increase and false negatives decrease. That is, the sensitivity increases. The disadvantage would then be obvious since the number of false positives also increase due to incorrect classification. As a result, specificity decreases.

In addition to sensitivity and specificity, the performance of a binary classification test can be measured with positive and negative predictive values. These are possibly more intuitively clear: the positive prediction value (PPV) answers the question “how likely it that there is a positive result, given that the SVM indicates a positive result?”. It is the proportion of true positives out of all positive results and is calculated using Equation 3.21.

\[
PPV = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}}
\]

The negative prediction value (NPV) is the same, but for negatives, and is calculated using Equation 3.22.
One should note, though, one important difference between the two concepts. That is, sensitivity and specificity are independent from the population in the sense that they don’t change depending on what the proportion of positives and negatives tested are. Indeed, you can determine the sensitivity of the test by testing only positive cases. However, the prediction values are dependent on the population.

As an example, classification techniques are often used in medical testing and say that you have a test for a disease with 99% sensitivity and 99% specificity. Say you test 2000 people, and 1000 of them are sick and 1000 of them are healthy. You are likely to get about 990 true positives, 990 true negatives, and 10 of false positives and negatives each. The positive and negative prediction values would be 99%, so the people can be quite confident about the result.

Say, however, that of the 2000 people only 100 are really sick. Now you are likely to get 99 true positives, 1 false negative, 1881 true negatives and 19 false positives. Of the 19+99 people tested positive, only 99 really have the disease - that means, intuitively, that given that your test result is positive, there’s only 84% chance that you really have the disease. On the other hand, given that your test result is negative, you can really be reassured: there’s only 1 chance in 1881, or 0.05% probability, that you have the disease despite of your test result [45].

3.7.2 Receiver Operating Characteristic Curve

In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented equivalently by plotting the fraction of true positives (TPR = true positive rate) vs. the fraction of false
To draw an ROC curve, only the true positive rate (TPR) and false positive rate (FPR) are needed. TPR determines a classifier or a diagnostic test performance on classifying positive instances correctly among all positive samples available during the test. FPR, on the other hand, defines how many incorrect positive results while they are actually negative among all negative samples available during the test. An example of a ROC graph is in Figure 3.9.

An ROC space is defined by FPR and TPR as $x$ and $y$ axes respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). Since TPR is equivalent with sensitivity and FPR is equal to $1 -$ specificity, the ROC graph is sometimes called the sensitivity vs $(1 -$ specificity) plot. Each prediction result or one instance of a confusion matrix represents one point in the ROC space.

The best possible prediction method would yield a point in the upper left corner or coordinate $(0,1)$ of the ROC space, representing 100% sensitivity (all true positives are
found) and 100% specificity (no false positives are found). The (0,1) point is also called a perfect classification. A completely random guess would give a point along a diagonal line (the so-called line of no-discrimination) from the left bottom to the top right corners. An intuitive example of random guessing is a decision by flipping coins (head or tail).

The diagonal line divides the ROC space in areas of good or bad classification/diagnostic. Points above the diagonal line indicate good classification results, while points below the line indicate wrong results (although the prediction method can be simply inverted to get points above the line) [46].
Chapter 4: Philadelphia Case Study

4.1 Philadelphia International Airport

FAA officials, airlines, air traffic controllers and others say Philadelphia plays a major role in delays up and down the coast thanks to poor airport design, bad weather, heavy traffic and close proximity to New York. Through September 2007, 68% of departures were on time in Philadelphia, better only than New York’s JFK International, Chicago’s O’Hare International and Liberty International in Newark, N.J. Fewer than two-thirds of arrivals were on time in Philadelphia during that period. The FAA has deemed Philadelphia a “pacing” airport that, because it sits in the middle of the busy East Coast air corridor, causes delays nationwide. It is debating how to improve the airport, which last year ranked 16th in the nation by passenger volume, but is consistently near the bottom of the 32 largest airports in on-time performance [47]. Because of these facts, Philadelphia was chosen to evaluate the dual formulation and alternate classification methods.

4.1.1 Philadelphia Airport Layout

Figure 4.1 shows that the primary runways at Philadelphia are 9L/R and 27 L/R. Poor runway arrangement limits the number of planes that can take off from the airport at once, especially during bad weather. Although a small runway was added in 1999, most of the layout dates back to the 1970s or earlier [47].

4.2 Linear Support Vector Machine Method

The SVM only classifies the data for a given AAR value. In order to create a useful tool, several SVM operations had to be done for one airport. The first step in the process was to
find the average demand rates for each hour during the day. These peaks are highlighted in Figure 4.2.

Figure 4.2 shows six peaks, but to reduce the dimensionality of the problem we chose only 0800, 1200, 1600, and 1800. For those time periods the most common AAR was 52, indicating normal operations, which occurred 60% of the time. During times of irregular operations the AARs are reduced to 48 for 20% of the time and 36 9% of the time. Because these three AARs constitute 89% of the possible AARs, these were set as the only possible solutions that the model will predict. Two SVMs were solved for each time period. The first SVM will classify whether or not the day had an AAR less than 52 or greater than or equal to 52. If the SVM classifies a given day and time period as greater than or equal to 52, then the tool will show a AAR of 52. This AAR would also indicate no GDP during this period. If the SVM predicts less than 52, then we would develop an SVM to test to see if the given day and time period is less than 48 or greater than or equal to 48. Again, if the SVM indicates greater than 48, then the tool sets the AAR to 48 and indicates a GDP. If the SVM indicates less than 48, then we set the AAR is set to 36 and also a GDP.
is predicted during this period. The duration of a predicted GDP is based on what time periods have GDP AARs.

4.2.1 Philadelphia Results

The SVM produced prediction vectors for Philadelphia, as well as all other prediction vectors can be found in that can be found in Appendix I. To evaluate how the SVM worked for Philadelphia, two methods were applied. The first method observed the success rate of the SVM prediction functions for the two test points for each time period. Data was also separated between training data, which was the data from January 2002 through December 2006, and testing data which is data from January 2007 through June 2007. The results for the training data are found in Table 4.1 and the results for the testing data are found in Table 4.2.

Table 4.1 and Table 4.2 indicate that the SVM algorithm was correct 81% of the time for the training data and 83% for the testing data. To create a meaningful tool containing these algorithms a set of rules was established to estimate the AAR. The tool only considers
### Table 4.1: Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.38</td>
<td>1.00</td>
<td>1.00</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.61</td>
<td>0.90</td>
<td>0.79</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.35</td>
<td>0.96</td>
<td>0.64</td>
<td>0.88</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.50</td>
<td>0.91</td>
<td>0.79</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.31</td>
<td>0.98</td>
<td>0.75</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.49</td>
<td>0.91</td>
<td>0.74</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.32</td>
<td>0.98</td>
<td>0.75</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.48</td>
<td>0.90</td>
<td>0.72</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.46</td>
<td>0.95</td>
<td>0.77</td>
<td>0.82</td>
<td>0.81</td>
</tr>
</tbody>
</table>

### Table 4.2: Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.40</td>
<td>1.00</td>
<td>1.00</td>
<td>0.83</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.49</td>
<td>0.94</td>
<td>0.81</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.44</td>
<td>0.99</td>
<td>0.85</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.40</td>
<td>0.92</td>
<td>0.70</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.36</td>
<td>0.98</td>
<td>0.77</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.39</td>
<td>0.92</td>
<td>0.66</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.28</td>
<td>0.98</td>
<td>0.73</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.35</td>
<td>0.92</td>
<td>0.61</td>
<td>0.79</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.40</td>
<td>0.96</td>
<td>0.75</td>
<td>0.84</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Table 4.3: Tool Results for Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>72%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>41%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>78%</td>
<td>70%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>64%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>46%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>74%</td>
<td>74%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>75%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>40%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>76%</td>
<td>76%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>75%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>39%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>75%</td>
<td>76%</td>
</tr>
</tbody>
</table>

three possible AARs, one associated with normal operations, one associated with a slight reduction in capacity, and one associated with a large reduction of capacity.

The first rule tested whether or not the point, that represents a day, was below 48. If it was below 48, then the AAR was determined based on a weighted average of the observed AARs below the tested rate, which for all four time periods was 36. If the SVM indicated the point was equal to or greater than 48 or less than 52, then we assumed the AAR was 48. All other results were assumed to be 52. Table 4.3 and Table 4.4 show the tool performance for the training and testing data.

Table 4.3 and Table 4.4 show that the accuracy is better at the extreme points then the points in the middle. This shows that the SVM method is better at finding extreme points on the edge instead of points inside.

4.2.2 Delay Prediction

Within the airline industry and air traffic management the AAR determines the airport capacity and is used to highlight the severity of the of a GDP, therefore it is the preferred
Table 4.4: Tool Results for Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>76%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>25%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>79%</td>
<td>80%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>92%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>30%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>77%</td>
<td>82%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>77%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>29%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>83%</td>
<td>84%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>73%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>28%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>81%</td>
<td>84%</td>
</tr>
</tbody>
</table>

prediction variable. Most flying consumers do not understand what AARs are and prefer to know what are the potential delays. The Weather Channel provides Airport Impact Maps like the one shown in Figure 4.3. The Weather Channel uses a Red, Amber, or Green rating system to highlight the airport impact. Although, the website does not explain the rating system, one would assume that Red impact means the most delays and Green impact means little to no delays. Amber is somewhere in the middle. To make the Weather Delay Prediction Tool applicable to the traveler, a delay prediction needed to be added.

Delay Prediction Method

A queuing system is often used to model a system where “customers”, in this case airplanes, must wait to be served or landed at a runway. Recall that the M/M/1 queuing system has exponential interarrival times and a single server with exponential service times. This seems to be idea to model landing planes, however, the demand rate is rarely constant as most airports have peak times. During these peak times, the arrival rate is larger than the service rate [48]. This causes the queue waiting time to go to infinity. Delays are bad, but not
that bad. To predict delays a more complex queuing system is required which is beyond the scope of this study, therefore a simpler method may be more appropriate.

A simple linear regression with AAR as the independent variable and delay as the dependent variable seems to be the next appropriate method to analyze. A regression was done for each time period and the $r^2$ was calculated. The $r^2$ is often interpreted as the proportion of response variation “explained” by the regressors in the model. Thus, $r^2 = 1$ indicates that the fitted model explains all variability in $y$, while $r^2 = 0$ indicates no “linear” relationship between the response variable and regressors. In this case, none of the $r^2$ results were greater than 0.42. An $r^2 = 0.42$ may be interpreted as follows: Approximately 42% of the variation in the response variable can be explained by the explanatory variable. The remaining 58% can be explained by unknown, lurking variables or inherent variability. Because of these results, regression was eliminated as a model to predict delay.

After eliminating more complex mathematical methods, it was determined that a simpler method may be more appropriate. Since the SVM model only predicts three potential AAR
outcomes, then the average delays during the time of those AARs would provide not only a mean value, but also a range. Using the AAR data from ASPM and the delay data from Bureau of Transportation Statistics website, the average delay was calculated for each corresponding AAR. To provide a range of values, the standard deviation was calculated and added and subtracted to the mean value to provide a range.

**Delay Results**

Delay values are negative if the average arrival is early. Therefore, many of the average values are negative. If this was true then the value was changed to zero. Delay values and ranges are rounded to the nearest minute in Table 4.5. Table 4.5 indicates the delay mean and the high and low range for each time period and predicted AAR. For instance, for the 1800 time period, if the model predicts an AAR of 36, then the delay mean is 54 minutes with a range as high as 96 and as low as 13 minutes. Now the flying consumer has information that they can use to plan their travel day.

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>0800</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>19</td>
</tr>
<tr>
<td>1200</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>37</td>
</tr>
<tr>
<td>1600</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>48</td>
</tr>
<tr>
<td>1800</td>
<td>Low</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>96</td>
</tr>
</tbody>
</table>
4.3 Dual Formulation

In Section 3, the dual formulation of the SVM was introduced. To see if there was any improvement in results the homogeneous polynomial function and the Gaussian radial-based kernel were chosen to derive prediction functions. The homogeneous polynomial function was chosen because there was a good chance that the data would be separated better by a curved hyperplane rather than a linear one and the Gaussian kernel was chosen since it is widely used [44]. These functions can be seen in Table 3.4. As seen in the primal formulation below, the given weather data is found in the constraint portion of the quadratic program.

\[
\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \\
\text{s.t. } y_i (w^T x + b) \geq 1 - \xi_i \\
\xi_i \geq 0, i = 1 \ldots N
\]

In this case it meant that there were 1826 constraints which should make the problem more difficult to solve. The assumed advantage of the dual formulation found in Equation 3.12 through Equation 3.14 is that it is easier to solve because, with the exception of one equality, all constraints are simple upper and lower bounds [44]. The problem with the dual formulation lies in the matrix found in the objective function. Since there are five years of data which result in 1826 data points, then if all of the data points are used, the result of \( YX^TXY \) is a 1826 \( \times \) 1826 matrix. To use AMPL to solve for the support vectors AMPL data file had to be loaded with the 1826 \( \times \) 1826 matrix and a 1826 \( \times \) 2 matrix which takes up over 39,000 KB while the file for the primal only takes up 240 KB. The NEOS file rejected this file as too large. The method also required preprocessing using MATLAB, found in Appendix B and Appendix C, to include the kernel function in the formulation. This was time consuming since the MATLAB results could not be directly inputed into the required AMPL data format and required extensive cutting and pasting.
Since the NEOS solver would not accept the data file, the number of data points was reduced to 500. These 500 points were randomly selected to ensure there was a wide array of data points to account for yearly and seasonal differences. Using the PHL data, the set of support vectors were determined using a homogeneous polynomial with $C = 0.1$ and $C = 1$ and the set of support vectors using a Gaussian kernel function with $C = 1$ and $C = 10$. The higher the $C$, the more classification errors are penalized in the objective function. A low $C$ may produce higher classification errors, but a high $C$ may create a set of predictor functions that overfits the data. Overfit data excels at classifying the training data, however, it does not perform well on testing data and hence may have little value as a predictor.

### 4.3.1 Dual Formulation Results

To compare the linear to the nonlinear kernel function, the $\alpha$ vector was computed using the dual formulation for each function and $C$ value. Only 500 of the data values were used, but in order to evaluate the results compared to the linear model, all 1826 data points were used to compare the function’s effectiveness in classifying the training data. From the set of the 500 vectors, each vector with a corresponding positive $\alpha_i$ is a support vector. Of the support vectors a smaller subset had $\alpha_i < C$. These vectors were used to compute $b$ using Equation 3.17. Using the support vectors and $b$ we then computed the classification using Equation 3.18.

If the function was greater than 0, then the data was classified as +1, and if the data was less than 0, then the data was classified as +1. The classifications were then compared to the training data and the results can be found in Table 4.6. Table 4.6 indicates that the linear model, the polynomial model with $C = 0.1$, and the Gaussian model with $C = 1$ all have similar performance with percent correct values of between 80% and 81%. At this point it is valuable to look at a ROC curve in Figure 4.4 to compare all of the values.

The Gaussian function with the $C = 10$ and the polynomial function with the $C = 1$ have points closest to the diagonal line which indicates that they have the poorest performance.
### Table 4.6: Predictions compared to training data

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>0.46</td>
<td>0.95</td>
<td>0.77</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>Polynomial C = 0.1</td>
<td>0.54</td>
<td>0.90</td>
<td>0.67</td>
<td>0.84</td>
<td>0.80</td>
</tr>
<tr>
<td>Polynomial C = 1</td>
<td>0.50</td>
<td>0.71</td>
<td>0.40</td>
<td>0.79</td>
<td>0.65</td>
</tr>
<tr>
<td>Gaussian C = 1</td>
<td>0.50</td>
<td>0.93</td>
<td>0.72</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>Gaussian C = 10</td>
<td>0.35</td>
<td>0.84</td>
<td>0.45</td>
<td>0.78</td>
<td>0.71</td>
</tr>
</tbody>
</table>

#### Figure 4.4: PHL Training Data Kernel ROC Graph
Table 4.7: Predictions compared to testing data

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>0.53</td>
<td>0.96</td>
<td>0.82</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Polynomial C = 0.1</td>
<td>0.40</td>
<td>0.90</td>
<td>0.64</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>Polynomial C = 1</td>
<td>0.41</td>
<td>0.66</td>
<td>0.33</td>
<td>0.73</td>
<td>0.58</td>
</tr>
<tr>
<td>Gaussian C = 1</td>
<td>0.46</td>
<td>0.87</td>
<td>0.54</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>Gaussian C = 10</td>
<td>0.38</td>
<td>0.82</td>
<td>0.39</td>
<td>0.80</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Both of these function and $C$ pairs have almost perfect performance when we only used 500 data points by correctly classifying over 94% of the points. However, when we expanded to 1826 data points, the performance worsened. This is probably a case of overfitting, therefore it is concluded that lower $C$ values should be used.

To further compare the proposed methods, testing data from the first six months of 2007 was used. Testing data is not used in computing the predictor function in the linear model or the support vectors in the nonlinear model. The results of this comparison can be found in Table 4.7. From Table 4.7 you can see that the linear model has a much better percent correct value than any of the other methods and $C$ values. Figure 4.5 compares the data with a ROC graph. The ROC Graph in Figure 4.5 further confirms that the linear model is more effective in producing AAR predictions since all of the blue squares are farther away from the red line and closer to the upper left hand corner. From this it is concluded that the other dual formulation overfit the data and therefore does a poor job of predicting future AARs from test data.

4.3.2 Nonlinear Conclusions

The initial results show that a linear model should be used to determine AARs. The nonlinear models require more preprocessing, more storage capacity, and in the end tend to overfit the data, which indicate that it is not an effective way of predicting future events. The linear model produces a linear predictor function which is easy to integrate into a simple spreadsheet and determine future AARs. Due to the support vectors required in
the nonlinear method required at least fifty times the calculations to produce an AAR estimate. Also if more historical data points are available, then the nonlinear method forces an exponential growth in storage requirements due the matrix required in the objective function. For instance, if another year is added, the number of data points is increased by 365 days. The linear model goes from storing a $1826 \times 57$ matrix to a $2191 \times 57$ matrix, but for the nonlinear model storage goes from a $1826 \times 1826$ matrix to a $2191 \times 2191$ matrix. The nonlinear storage requirements would increase by over 40%.

4.4 Evaluation of Alternate Classification Methods

The previous section evaluated the utility of expanding it beyond the linear hyperplane to an non-linear hyperplane. This section goes beyond the support vector machine and examines other supervised learning methods that may classify data more effectively. In the first part of the section, a linear least square regression model is applied to the Philadelphia data to
see if the performance is better than the SVM. Least squares regression was chosen since it is a very commonly used predictive tool[48]. The next part of the chapter examines the k-nearest neighbor algorithm which is one of the simplest classification methods available. This was chosen to determine whether or not a simpler procedure may in fact perform better. Finally, the last part used recursive partitioning to create a decision tree. This was chosen because of the complexity to see if a more complex technique will provide better results. All techniques used the Philadelphia data in order to provide a common dataset to compare results.

4.5 Least Squares Regression

Least squares regression assumes that there is a linear relationship between a dependent variable, in this case the set of actual AARs, and a set of independent variables, which are represented by the TAF data. The intent of using linear regression in this chapter is not to develop an equation and then test the goodness of fit, but instead to use regression to predict future AARs and test using the same evaluation methods as the classification models. To develop the prediction equation, a least squares non-linear program, as seen in Appendix D, is coded into AMPL.

4.5.1 Least Squares Method

The AMPL code produced an $m$ vector and an $b$ $y$-intercept to be used in the typical $y = mx + b$ equation. These results are found in Appendix E. To make it more like a classification method, the actual predict $y$ value was not used, but instead $y$ was compared to the same break values as the other classification methods. If the $y$ value was less than 48, then it was assumed that the point AAR was 36. If the $y$ value is greater than or equal to 48 and less than 52, then the predicted value is 48. Finally, if the $y$ value is greater than or equal to 52, then the AAR is assumed to be 52. This allows the algorithm to be evaluated the same way as the other classification models.
Table 4.8: Least Squares Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.77</td>
<td>0.78</td>
<td>0.46</td>
<td>0.93</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.98</td>
<td>0.13</td>
<td>0.42</td>
<td>0.90</td>
<td>0.46</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.73</td>
<td>0.81</td>
<td>0.43</td>
<td>0.94</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.96</td>
<td>0.13</td>
<td>0.42</td>
<td>0.83</td>
<td>0.46</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.65</td>
<td>0.84</td>
<td>0.42</td>
<td>0.93</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.96</td>
<td>0.16</td>
<td>0.39</td>
<td>0.87</td>
<td>0.45</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.66</td>
<td>0.84</td>
<td>0.43</td>
<td>0.93</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.95</td>
<td>0.15</td>
<td>0.39</td>
<td>0.85</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.88</td>
<td>0.53</td>
<td>0.41</td>
<td>0.93</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 4.9: Least Squares Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.84</td>
<td>0.77</td>
<td>0.49</td>
<td>0.95</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>1.00</td>
<td>0.01</td>
<td>0.33</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.80</td>
<td>0.78</td>
<td>0.36</td>
<td>0.96</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>1.00</td>
<td>0.01</td>
<td>0.32</td>
<td>1.00</td>
<td>0.32</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.68</td>
<td>0.88</td>
<td>0.50</td>
<td>0.94</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>1.00</td>
<td>0.05</td>
<td>0.28</td>
<td>1.00</td>
<td>0.30</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.66</td>
<td>0.88</td>
<td>0.50</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.98</td>
<td>0.03</td>
<td>0.27</td>
<td>0.80</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.91</td>
<td>0.46</td>
<td>0.33</td>
<td>0.94</td>
<td>0.56</td>
</tr>
</tbody>
</table>

4.5.2 Least Squares Result

Table 4.8 and Table 4.9 shows the performance of linear regression.

Table 4.8 and Table 4.9 indicate that the least square regression algorithm was correct only 63% of the time for the training data and only 56% for the testing data. The performance of the least squares algorithm is poor compared to the SVM. The results of the testing data can be found in Table 4.10 and Table 4.11.

In Table 4.10 and Table 4.11, it initially appears that the algorithm effectively predicts an AAR of 52. However, upon further review, the algorithm does very poorly when predicting an AAR of 48. Most of the data points fall within the 48 prediction, with none higher than
Table 4.10: Tool Results for Least Squares Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>46%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>17%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>90%</td>
<td>61%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>43%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>21%</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>83%</td>
<td>60%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>42%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>19%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>87%</td>
<td>64%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>43%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>19%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>85%</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 4.11: Tool Results for Least Squares Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>49%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>8%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>100%</td>
<td>67%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>36%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>15%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>100%</td>
<td>69%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>50%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>10%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>100%</td>
<td>73%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>50%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>80%</td>
<td>73%</td>
</tr>
</tbody>
</table>
20% in accuracy. This shows that most of the predictions are not correct.

4.6 k-Nearest Neighbor Algorithm

The $k$-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its $k$ nearest neighbors. $k$ is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose $k$ to be an odd number as this avoids tied votes. The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. Typically the Euclidean distance is used, though other distance measures, such as the Manhattan distance could in principle be used instead [49].

4.6.1 Nearest Neighbor Method

The first step in the $k$-nearest neighbor algorithm was to use the MATLAB code found in Appendix F to find the distance from each test point to everyone of the training points. This produced a distance matrix for all 181 of the test points. The next step was to find the five lowest distances. This provided the five nearest neighbors to the test points. Then, based on the training data, the five nearest neighbor points were classified as 1 if it is within the set or -1 if it is not within the set. The classification values were summed and if the value was positive then the test point was classified as a 1 and if the sum was negative then the test point was classified as -1.

4.6.2 Nearest Neighbor Results

The $k$-nearest neighbor algorithm only uses the training data as a set of data from comparison, so unlike the other methods, the performance can only be evaluated with Testing data.
Table 4.12: Philadelphia Nearest Neighbor Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>48</td>
<td>0.37</td>
<td>0.94</td>
<td>0.61</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.46</td>
<td>0.85</td>
<td>0.60</td>
<td>0.76</td>
<td>0.72</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.24</td>
<td>0.97</td>
<td>0.55</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.46</td>
<td>0.83</td>
<td>0.55</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.25</td>
<td>0.98</td>
<td>0.70</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.33</td>
<td>0.86</td>
<td>0.46</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.21</td>
<td>0.98</td>
<td>0.67</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.41</td>
<td>0.83</td>
<td>0.47</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>Combined</td>
<td>0.37</td>
<td>0.91</td>
<td>0.55</td>
<td>0.83</td>
<td>0.78</td>
<td></td>
</tr>
</tbody>
</table>

The algorithm performance is found in Table 4.12. As seen in Table 4.12, the \(k\)-nearest neighbor algorithm performs quite well. Table 4.13 shows how the algorithm would work if it was entered into a tool to predict AARs. Table 4.13 shows, that like other methods, the tool is more accurate at extreme points then points in the middle.

4.7 Recursive Partitioning

Recursive partitioning is a statistical method used for classification. Recursive partitioning creates a decision tree that strives to correctly classify members of the population based on a dichotomous dependent variable [50]. As compared to other classification techniques that create a formula used to calculate whether or not the data is in one data set or the other, recursive partition creates a rule such as, “if a is x, y, or z then the point is in a.”

Recursive partitioning methods create more intuitive models that do not require the user to perform calculations [51]. They allow varying prioritizing of misclassifications in order to create a decision rule that has more sensitivity or specificity and they may be more accurate [52]. However, recursive partitioning must bin continuous variables and may overfit data [53].
Table 4.13: Tool Results for Nearest Neighbor Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>70%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>9%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>79%</td>
<td>67%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>55%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>22%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>78%</td>
<td>69%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>80%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>24%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>80%</td>
<td>173%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>78%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>80%</td>
<td>73%</td>
</tr>
</tbody>
</table>

4.7.1 Recursive Partitioning Method

Recursive partitioning is the strategy that generates tree-structured prediction models. A modeling method using the recursive-partitioning strategy evaluates at each step different ways of partitioning the training data into subsets (e.g., by considering possible values for an attribute), and chooses the best partitioning according to some criterion (e.g., the alternative that minimizes the variance of the prediction variable), and continues recursively with each remaining subset until some conditions hold (e.g., all examples in the subset belong to the same class). It results in a hierarchically organized set of rules (a tree model) that can be represented as a decision tree [54].

4.7.2 Recursive Partitioning Results

All of the decision trees were determined using Rule Discovery System developed by Compu- mine. The Rule Discovery System created assessed rule-based prediction models from existing data which can be found in Appendix I. These rules were programmed into an Excel spreadsheet. With all of the training and testing data also in the spreadsheet, the rules
Table 4.14: Recursive Partitioning Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.42</td>
<td>0.94</td>
<td>0.61</td>
<td>0.87</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.62</td>
<td>0.87</td>
<td>0.76</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.32</td>
<td>0.96</td>
<td>0.61</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.47</td>
<td>0.88</td>
<td>0.76</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.28</td>
<td>0.98</td>
<td>0.69</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.48</td>
<td>0.81</td>
<td>0.37</td>
<td>0.87</td>
<td>0.75</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.23</td>
<td>0.98</td>
<td>0.72</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.50</td>
<td>0.89</td>
<td>0.74</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.46</td>
<td>0.92</td>
<td>0.66</td>
<td>0.83</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 4.15: Recursive Partitioning Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>48</td>
<td>0.21</td>
<td>0.94</td>
<td>0.47</td>
<td>0.82</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.53</td>
<td>0.79</td>
<td>0.54</td>
<td>0.77</td>
<td>0.70</td>
</tr>
<tr>
<td>1200</td>
<td>48</td>
<td>0.26</td>
<td>0.97</td>
<td>0.58</td>
<td>0.88</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.37</td>
<td>0.82</td>
<td>0.56</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>1600</td>
<td>48</td>
<td>0.21</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.46</td>
<td>0.87</td>
<td>0.58</td>
<td>0.81</td>
<td>0.76</td>
</tr>
<tr>
<td>1800</td>
<td>48</td>
<td>0.10</td>
<td>0.99</td>
<td>0.75</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.41</td>
<td>0.85</td>
<td>0.50</td>
<td>0.79</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.35</td>
<td>0.91</td>
<td>0.56</td>
<td>0.82</td>
<td>0.78</td>
</tr>
</tbody>
</table>

predicted the outcome. These outcomes were then evaluated using the same method that was used for the SVM method. The results for the training data are found in Table 4.14 and the results for the testing data are found in Table 4.15.

Table 4.14 and Table 4.15 indicate that the SVM algorithm was correct 80% of the time for the training data and 78% for the testing data. These statistics were not as good as the statistics for the SVM method, so the method was further investigated to see if the algorithm perform better in a tool. The results can be found in Table 4.16 and Table 4.17.

Like all the other methods investigated, Table 4.16 and Table 4.17 shows that the accuracy is better at the extreme points then the points in the middle.
Table 4.16: Tool Results for Recursive Partitioning Philadelphia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>61%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>79%</td>
<td>61%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>61%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>41%</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>76%</td>
<td>60%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>69%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>41%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>77%</td>
<td>64%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>72%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>41%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>77%</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 4.17: Tool Results for Recursive Partitioning Philadelphia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>36</td>
<td>47%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>79%</td>
<td>67%</td>
</tr>
<tr>
<td>1200</td>
<td>36</td>
<td>67%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>34%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>80%</td>
<td>69%</td>
</tr>
<tr>
<td>1600</td>
<td>36</td>
<td>100%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>27%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>82%</td>
<td>73%</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>75%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>17%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>82%</td>
<td>73%</td>
</tr>
</tbody>
</table>
4.8 Alternate Method Conclusions

The poor performance of the linear regression algorithm is highlighted in Table 4.8, Table 4.9, Table 4.10, and Table 4.11. Extremely poor accuracy rates for the middle 48 value can be overlooked in the other two methods since most of the data points fall in one of the extreme points. For linear regression most of the data points fall within the 48 bin and have very low accuracies. Therefore, linear regression should be excluded from any further consideration.

The tables for Nearest Neighbor Method and the Recursive Partitioning Method all seem to compare favorable to the SVM, therefore another comparison method must be used to draw any conclusion to what is more suitable. The best way to compare the Recursive Partitioning Method, the Nearest Neighbor Method, and the SVM is to plot the true positive rate and the false positive rate on a ROC Chart. The Nearest Neighbor algorithm does not have any training data results, so the ROC graph in Figure 4.6 only contains the data from the SVM and Recursive Partition. You can see from Figure 4.6 that the points that represent the SVM are farther from the line of no discrimination and closer to the upper right hand corner then the point that represent the Recursive Partitioning Decision Tree.
Therefore, it is concluded from the graph, that the SVM more accurately predicts the training data. Figure 4.7 highlights the testing data and includes the data for the Nearest Neighbor algorithm. Figure 4.7 shows that SVM is also better than the other two methods. Therefore, the rest of this paper will add other airports to the analysis to see if the SVM can be applied to many different types of airports.
Chapter 5: Airport Results

Based on the analysis in the previous chapters, this paper has shown that the linear support vector machine is the best method to predict aircraft arrival rates, ground delay programs and delays. The next step was to apply this method to other major domestic airports. Besides Philadelphia, this paper examined Newark International Airport, Chicago O’Hare International Airport, Atlanta Hartsfield International Airport, LaGuardia Airport, John F. Kennedy International Airport, Reagan National Airport, and Dulles International Airport.

5.1 Newark International Airport

Newark International Airport is one of the three major New York area airports that is managed by the Port Authority of New York and New Jersey. Newark Liberty is the second-largest hub for Continental Airlines, which is the airport’s largest tenant (operating all of Terminal C and part of Terminal A). FedEx Express operates one of their ten major cargo hubs at Newark. In 2006, Newark Airport handled about 36 million passengers. Along with JFK’s 43 million and LaGuardia’s 26 million, New York’s airspace surpasses that of Chicago to become the busiest in the United States. Newark is the tenth busiest airport in the United States and the nation’s fifth busiest international air gateway; JFK ranks first [55].

5.1.1 Airport Layout

Newark Liberty International Airport covers 2,027 acres and has three runways and one helipad. As seen in Figure 5.1 Newark has three runways. Most departing traffic use Runway 4L/22R, while most arriving traffic use 04R/22L, and 11/29 is used more often when the crosswinds on the two main runways is strong enough.
Figure 5.1: Newark Airport Map
As seen in Figure 5.1, the main runways orient in the direction of 40 and 220 degrees, we assumed that a cross wind was and wind between 270 and 350 and between 90 and 170 degrees.

5.1.2 Newark Evaluation Method

A support vector machine is a classification method that divides data into two sets. This data requires a point to divide the data. In this case that point is from the aircraft arrival rates for the airport. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period. The Newark Airport demand chart in Figure 5.2 shows peaks at 0800, 1000, 1600, and 1900 local time, so they were chosen as the points to analyze. For each each day from January 2002 through December 2006, the AAR was recorded for each time period. Then for each time period the number of times a specific AAR was observed was counted. This count revealed that the most common AAR was 44 and the second most common AAR was 42 while the third most common AAR was 40. To divide the data, the AAR values of 41 and 43 were chosen as points to create a line to divide the data into two classes. One class is above the AAR and while the other class is equal to
<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>41</td>
<td>0.37</td>
<td>0.92</td>
<td>0.69</td>
<td>0.75</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.72</td>
<td>0.71</td>
<td>0.71</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td>1000</td>
<td>41</td>
<td>0.35</td>
<td>0.94</td>
<td>0.73</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.73</td>
<td>0.72</td>
<td>0.70</td>
<td>0.74</td>
<td>0.72</td>
</tr>
<tr>
<td>1600</td>
<td>41</td>
<td>0.39</td>
<td>0.92</td>
<td>0.74</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.52</td>
<td>0.83</td>
<td>0.74</td>
<td>0.64</td>
<td>0.67</td>
</tr>
<tr>
<td>1900</td>
<td>41</td>
<td>0.41</td>
<td>0.92</td>
<td>0.72</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.58</td>
<td>0.77</td>
<td>0.71</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>0.53</td>
<td>0.85</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>

or less than the AAR.

5.1.3 Newark Results

The SVM produced prediction vectors for Newark, as well as all other prediction vectors can be found in that can be found in Appendix I. To evaluate how the SVM worked for Newark, two methods were applied. The first method observed the success rate of the SVM prediction functions for the two test points for each time period. Data was also separated between training data, which was the data from January 2002 through December 2006, and testing data which is data from January 2007 through June 2007. The results for the training data are found in Table 5.1 and the results for the testing data are found in Table 5.2.

Table 5.1 and Table 5.2 indicate that the SVM algorithm was correct 72% of the time for the training data and 64% for the testing data. These results may not have been as good as Philadelphia, but one must consider that Newark is a congested airport and that congestion is probably responsible for some of the reduction in airport capacity. Therefore weather may not have as great of an influence on AARs as less congested airports like Philadelphia.

Table 5.1 and Table 5.2 describe the performance of the algorithm. To create a meaningful tool containing these algorithms a set of rules was established to estimate the AAR.
The tool only considers three possible AARs, one associated with normal operations, one associated with a slight reduction in capacity, and one associated with a large reduction of capacity.

The first rule tested whether or not the point, that represents a day, was below 41. If it was below 41, then the AAR was determined based on a weighted average of the observed AARs below the tested rate, which for all four time periods was 37. If the SVM indicated the point was equal to or greater than 41 or less than 43, then we assumed the AAR was 42. Any SVM result over 43 was assumed to be the weighted average of all observed AARs above 43. This value was different for the four chosen time periods. For 0800 and 1000 local time, the weighted average AAR was 45 while for 1600 and 1900 local time the weighted average was 47. To see how well the tool works, actual data was divided into three bins the same way as the rules developed for the tool. Table 5.3 and Table 5.4 show the tool results for the training data and the testing data. Table 5.5 shows the AAR and the corresponding delay per time period. This was calculated the same way as the Philadelphia delay seen in Chapter 4.

Like the results from Philadelphia, the Newark results have success at the extreme points, but lack certainty for the middle AAR value. With the successful accuracy values at the extreme points though, a manager can still make one of three conclusions;

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>41</td>
<td>0.43</td>
<td>0.95</td>
<td>0.70</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.10</td>
<td>0.99</td>
<td>0.91</td>
<td>0.45</td>
<td>0.48</td>
</tr>
<tr>
<td>1000</td>
<td>41</td>
<td>0.43</td>
<td>0.98</td>
<td>0.83</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.12</td>
<td>0.99</td>
<td>0.92</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td>1600</td>
<td>41</td>
<td>0.23</td>
<td>1.00</td>
<td>1.00</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.26</td>
<td>1.00</td>
<td>1.00</td>
<td>0.51</td>
<td>0.59</td>
</tr>
<tr>
<td>1900</td>
<td>41</td>
<td>0.24</td>
<td>0.98</td>
<td>0.91</td>
<td>0.59</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.20</td>
<td>1.00</td>
<td>1.00</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Combined</td>
<td>0.21</td>
<td>0.98</td>
<td>0.91</td>
<td>0.60</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.3: Tool Results for Newark Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>37</td>
<td>69%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>30%</td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>74%</td>
<td>48%</td>
</tr>
<tr>
<td>1000</td>
<td>37</td>
<td>73%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>29%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>75%</td>
<td>50%</td>
</tr>
<tr>
<td>1600</td>
<td>37</td>
<td>74%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>20%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>64%</td>
<td>65%</td>
</tr>
<tr>
<td>1900</td>
<td>37</td>
<td>72%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>24%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>67%</td>
<td>59%</td>
</tr>
</tbody>
</table>

Table 5.4: Tool Results for Newark Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>37</td>
<td>68%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>48%</td>
<td>89%</td>
</tr>
<tr>
<td>1000</td>
<td>37</td>
<td>83%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>49%</td>
<td>90%</td>
</tr>
<tr>
<td>1600</td>
<td>37</td>
<td>100%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>52%</td>
<td>85%</td>
</tr>
<tr>
<td>1900</td>
<td>37</td>
<td>91%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>43%</td>
<td>87%</td>
</tr>
</tbody>
</table>
Table 5.5: Newark Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>45</td>
</tr>
<tr>
<td>0800</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>1600</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
</tr>
</tbody>
</table>

- good chance of normal operations
- good chance of irregular operations due to weather
- with less certainty, there is a chance of reduced operations due to decreased capacity.

5.2 Chicago O’Hare International Airport

Chicago O’Hare International Airport is located 17 miles northwest of the Chicago loop. It is the largest hub of United Airlines and the second largest hub of American Airlines. In 2005, the airport had 972,246 aircraft operations, an average of 2,663 per day [56]. Prior to 2005, O’Hare was the world’s busiest airport in terms of takeoffs and landings. That year, mainly due to limits imposed by the federal government to reduce flight delays, Atlanta International Airport became the busiest [57]. In 2006, only 70% of flight leaving O’Hare had an on time departure [58].
5.2.1 Airport Layout

There are 6 primary runways, arranged tangentially in 3 pairs of parallel sets. All but one of O’Hare’s runways intersect, which can create problems in times of inclement weather, congestion at the airport, or wind. Because of this intersection, controllers must wait until a runway is free of traffic before they can clear another plane on an intersecting runway.

As seen in Figure 5.3, O’Hare has three sets of parallel runways that serve landings and takeoffs from most directions. Runways 4 L/R and 22 L/R are not as long then the other parallel set of runways, so it is assumed a crosswind was any landing that had to utilize those runways.

5.2.2 O’Hare Evaluation Method

The data was divided similar to the way previous data was divided using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the
given time period. The O’Hare Airport demand chart in Figure 5.4 shows peaks at 0800, 1000, 1200, and 1900 local time, so they were chosen as the points to analyze. The AAR count revealed that the most common AAR was 100 and the second most common AAR was 96 while the third most common AAR was 80. Initially the data was divided to use these three AAR values. When trying to separate 100 AAR data from the rest, the SVM produced a predictor vector $w$ with all zeros and an $b$ of -1, so the SVM was not effective in determining the differences from an AAR of 96 or an AAR of 100. Therefore the high AAR value was set to 100 and all AAR values greater than 95 were assumed to be 100. The fourth most common AAR value was 90, so data was classified either 80, 90, or 100.

### 5.2.3 O’Hare Results

Like Newark the method was evaluated by observing the success rate of the SVM prediction functions for the two test points for each time period. The training data and testing data
Table 5.6: O’Hare Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>90</td>
<td>0.33</td>
<td>0.95</td>
<td>0.69</td>
<td>0.82</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.31</td>
<td>0.93</td>
<td>0.69</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>1000</td>
<td>90</td>
<td>0.28</td>
<td>0.96</td>
<td>0.69</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.31</td>
<td>0.94</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>1200</td>
<td>90</td>
<td>0.30</td>
<td>0.96</td>
<td>0.66</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.25</td>
<td>0.95</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>1900</td>
<td>90</td>
<td>0.24</td>
<td>0.98</td>
<td>0.77</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.31</td>
<td>0.93</td>
<td>0.68</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.29</td>
<td>0.95</td>
<td>0.69</td>
<td>0.79</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 5.7: O’Hare Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>90</td>
<td>0.32</td>
<td>0.96</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.45</td>
<td>0.93</td>
<td>0.75</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>1000</td>
<td>90</td>
<td>0.48</td>
<td>0.96</td>
<td>0.78</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.46</td>
<td>0.95</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>1200</td>
<td>90</td>
<td>0.38</td>
<td>0.94</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.32</td>
<td>0.95</td>
<td>0.79</td>
<td>0.69</td>
<td>0.71</td>
</tr>
<tr>
<td>1900</td>
<td>90</td>
<td>0.18</td>
<td>0.99</td>
<td>0.91</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.25</td>
<td>0.94</td>
<td>0.72</td>
<td>0.65</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.35</td>
<td>0.95</td>
<td>0.78</td>
<td>0.75</td>
<td>0.76</td>
</tr>
</tbody>
</table>

used the same dates of January 2002 to December 2006 and January 2007 to June 2007. The results for the training data are found in Table 5.6 and the results for the testing data are found in Table 5.7.

Table 5.6 and Table 5.7 indicate that the SVM algorithm was correct 78% of the time for the training data and 76% for the testing data. These results are very encouraging because, although O’Hare is a very busy airport that is subjected to congestion, the weather data was still effective at predicting the AARs. After testing the performance of the algorithm, the next task was to test the performance of the tool.

The rules used for O’Hare are similar to the rules for Newark, except the test points
Table 5.8: Tool Results for O’Hare Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>80</td>
<td>69%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>21%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>72%</td>
<td>84%</td>
</tr>
<tr>
<td>1000</td>
<td>80</td>
<td>69%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>19%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>74%</td>
<td>86%</td>
</tr>
<tr>
<td>1200</td>
<td>80</td>
<td>66%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>21%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>74%</td>
<td>88%</td>
</tr>
<tr>
<td>1900</td>
<td>80</td>
<td>77%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>18%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>76%</td>
<td>86%</td>
</tr>
</tbody>
</table>

are different. The first rule tested whether or not the point, that represents a day, was below 90. If it was below 90, then the AAR was determined based on a weighted average of the observed AARs below the tested rate, which for all four time periods was 80. If the SVM indicated the point was equal to or greater than 90 or less than 96, then we assumed the AAR was 90. Any SVM result over 96 was assumed to be the weighted average of all observed AARs above 96. Unlike Newark, this AAR value, 100, was the same for all time periods. To see how well the tool works, actual data was divided into three bins the same way as the rules developed for the tool. Table 5.8 and Table 5.9 show the tool results for the training data and the testing data. Table 5.10 shows the potential delays based on the AAR and time period.

As with the other airports the tool experienced more success at the extreme points then at the middle points. This method would still be valuable since it predicts normal operations and irregular operations with at least 64% accuracy.
### Table 5.9: Tool Results for O'Hare Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>80</td>
<td>77%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>78%</td>
<td>80%</td>
</tr>
<tr>
<td>1000</td>
<td>80</td>
<td>78%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>80%</td>
<td>82%</td>
</tr>
<tr>
<td>1200</td>
<td>80</td>
<td>76%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>70%</td>
<td>83%</td>
</tr>
<tr>
<td>1900</td>
<td>80</td>
<td>91%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>6%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>66%</td>
<td>85%</td>
</tr>
</tbody>
</table>

### Table 5.10: O'Hare Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80  90  100</td>
</tr>
<tr>
<td>0800</td>
<td>Low</td>
<td>0   0   0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>5   0   0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>16  3   1</td>
</tr>
<tr>
<td>1000</td>
<td>Low</td>
<td>0   0   0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>22  0   0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>50  10  4</td>
</tr>
<tr>
<td>1200</td>
<td>Low</td>
<td>7   0   0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>43  5   0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>80  19  11</td>
</tr>
<tr>
<td>1900</td>
<td>Low</td>
<td>25  0   0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>75  17  9</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>126 44  28</td>
</tr>
</tbody>
</table>

90
5.3 Atlanta Hartsfield International Airport

Hartsfield-Jackson Atlanta International Airport is located seven miles south of the central business district of Atlanta. Hartsfield held its ranking as the world’s busiest airport in 2006, both in terms of passengers and number of flights, by accommodating 84.8 million passengers and 976,447 flights respectively [59].

In 1999 Hartsfield’s leadership developed a plan for multiple construction projects with the intention of preparing the airport to handle a projected demand of 121 million passengers in 2015. The first of the projects undertaken, the airport’s fifth runway, opened on May 27, 2006. The 9000 foot runway is the first runway added to the airport since 1984 and is expected to increase the capacity for landings and take-offs by 40%, from an average of 184 flights per hour to 237 flights per hour. Along with the construction of the fifth runway, the tallest control tower in the United States was built to see the entire length of the runway [60].

An “end-around taxiway”, officially named Taxiway Victor, which opened in April 2007, is expected to save an estimated $26 million to $30 million in fuel by allowing airplanes landing on the northernmost runway to taxi to the gate area without hindering other aircraft taking off. The taxiway drops approximately 30 feet from the runway elevation to allow takeoffs to continue [61]. The thirty-five year old runway 8R-26L was rehabilitated and reopened on November 4, 2006.

5.3.1 Airport Layout

A seen in Figure 5.5 there are 5 primary runways, arranged tangentially in 2 pairs of parallel sets and one single runway. All runways are in an east west orientation.

5.3.2 Atlanta Evaluation Method

The data was divided similar to the way we divided previous data using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period.
Figure 5.5: Atlanta Airport Map

[43]
The Atlanta Airport demand chart in Figure 5.6 shows peaks at 0800, 1100, 1600, and 2000 local time, so they were chosen as the points to analyze. As described above, Atlanta’s airport has had extensive improvements during the time that the training data was collected. Because of this, the training data had to be normalized so that the airport capacity as expressed as the AAR was consistent for all dates within the training set. Before the fifth runway was completed, the standard AAR for the airport was 96. After the runway’s completion, the standard AAR increased to 120. In late August of 2006, the rehabilitation of runway 8R and 26L began. This reduced the standard normal AAR to 96 again until the work completion in November of 2006 when the AAR increased to 126. All values were normalized to 126 through a simple ration of the normal AAR of the given date to the present normal AAR of 126.

Because of the normalization of the data for Atlanta, it was more difficult to count the number of AAR occurrences, so instead the method searched for a break point in the data that would create three separate data bins. Once this was done, then the weighted average
between the break point provided AAR values. Figure 5.7 shows that the data is divided at 124 and 115. These points were chosen as break points for the data. Between these points weighted averages were calculated and found to be a high of 128, a reduced operations value of 120 and a irregular operations value of 104. If the algorithm indicated less than 115, then the tool predicted 104. If the algorithm indicated greater than or equal to 115 and less than 124, then the tool predicted 120 and if the algorithm indicated greater than or equal to 124, then the tool predicted an AAR of 128.

5.3.3 Atlanta Results

The results for the Atlanta training data are found in Table 5.12 and the results for the Atlanta testing data are found in Table 5.13.

Table 5.11 and Table 5.12 indicate that the SVM algorithm was correct 74% of the time for the training data and 63% for the testing data. The training data for Atlanta was different from the previous data sets because of the requirement to normalize data due to the airport improvements. This might explain the low performance for the test data. Also, with Atlanta being the busiest airport in the United States, other congestion issues may reduce the capacity and weather may not have as much of an impact. However, the results
### Table 5.11: Atlanta Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>115</td>
<td>0.61</td>
<td>0.86</td>
<td>0.74</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.83</td>
<td>0.60</td>
<td>0.80</td>
<td>0.65</td>
<td>0.75</td>
</tr>
<tr>
<td>1100</td>
<td>115</td>
<td>0.57</td>
<td>0.86</td>
<td>0.72</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.82</td>
<td>0.61</td>
<td>0.80</td>
<td>0.64</td>
<td>0.75</td>
</tr>
<tr>
<td>1600</td>
<td>115</td>
<td>0.54</td>
<td>0.85</td>
<td>0.69</td>
<td>0.75</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.99</td>
<td>0.06</td>
<td>0.74</td>
<td>0.67</td>
<td>0.74</td>
</tr>
<tr>
<td>2000</td>
<td>115</td>
<td>0.25</td>
<td>0.95</td>
<td>0.68</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.99</td>
<td>0.04</td>
<td>0.70</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.77</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

### Table 5.12: Atlanta Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>115</td>
<td>0.25</td>
<td>0.92</td>
<td>0.67</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.79</td>
<td>0.51</td>
<td>0.69</td>
<td>0.63</td>
<td>0.67</td>
</tr>
<tr>
<td>1100</td>
<td>115</td>
<td>0.26</td>
<td>0.93</td>
<td>0.75</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.66</td>
<td>0.51</td>
<td>0.70</td>
<td>0.45</td>
<td>0.60</td>
</tr>
<tr>
<td>1600</td>
<td>115</td>
<td>0.36</td>
<td>0.89</td>
<td>0.68</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.98</td>
<td>0.04</td>
<td>0.57</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td>2000</td>
<td>115</td>
<td>0.12</td>
<td>0.98</td>
<td>0.80</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.98</td>
<td>0.04</td>
<td>0.57</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.60</td>
<td>0.66</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Table 5.13: Tool Results for Atlanta Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>104</td>
<td>74%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>36%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>65%</td>
<td>31%</td>
</tr>
<tr>
<td>1100</td>
<td>104</td>
<td>72%</td>
<td>31%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>37%</td>
<td>37%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>64%</td>
<td>32%</td>
</tr>
<tr>
<td>1600</td>
<td>104</td>
<td>69%</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>40%</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>67%</td>
<td>2%</td>
</tr>
<tr>
<td>2000</td>
<td>104</td>
<td>68%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>42%</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>59%</td>
<td>2%</td>
</tr>
</tbody>
</table>

are still consistent enough to warrant integration into the tool.

The rules for making prediction were discussed earlier in this chapter. Table 5.13 and Table 5.14 show how well the algorithm performs when acting as a prediction tool. Table 5.15 shows the delay predictions for Atlanta.

Based on the training data and the testing data the tool does an effective job at predicting irregular operations caused by weather with a high accuracy rate at predicting an AAR of 104. However, neither the training data or the testing data predict an AAR of 128 very often. Again, this may be attributed to how busy the airport is or the fact that the training data had to be normalized.

5.4 LaGuardia Airport

LaGuardia is the smallest of the New York area’s three primary commercial airports. Most flights from LaGuardia go to destinations within the US and Canada, as well as service to Aruba, the Bahamas and Bermuda. A perimeter rule prohibits incoming and outgoing flights that exceed 1,500 miles except on Saturdays, when the ban is lifted, and to Denver,
Table 5.14: Tool Results for Atlanta Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>104</td>
<td>67%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>22%</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>63%</td>
<td>33%</td>
</tr>
<tr>
<td>1100</td>
<td>104</td>
<td>76%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>17%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>46%</td>
<td>40%</td>
</tr>
<tr>
<td>1600</td>
<td>104</td>
<td>68%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>18%</td>
<td>77%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>60%</td>
<td>3%</td>
</tr>
<tr>
<td>2000</td>
<td>104</td>
<td>80%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>20%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>60%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 5.15: Atlanta Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
<th>104</th>
<th>120</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>Low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>Low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>23</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td>Low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>19</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>38</td>
<td>14</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>Low</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>37</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>73</td>
<td>31</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>
which was grandfathered in so most transcontinental and international flights use the area’s other two major airports, JFK and Newark [62].

LaGuardia airport has a long history of severe congestion. Beginning in 1968, the FAA instituted a High-Density-Rule (HDR) that limited the number of takeoff and landings per hour or half hour during certain times during the day. In late 2000, the FAA provided an additional 300 slots into LGA to new entrant carriers or to airlines that fly 70 seat or less aircraft to small or non-hub airports. The results was a significant increase in delays. Currently, the growth in air traffic has put substantial pressure on the airports infrastructure, which has limited possibilities for expansion. Average arrival delay per flight reached 28 minutes in July 2000 and 31 min in July 2005 [63].

5.4.1 Airport Layout

As seen in Figure 5.8 there is only a set of two intersecting runways. When aircraft are arriving on Runway 22 and departing on Runway 31, controllers frequently need to apply additional separation between aircraft due to wake vortex turbulence. This situation requires controllers to apply 2 minutes of separation between subsequent Runway 22 arrivals and Runway 31 departures, almost double the separation if wake turbulence were not a factor. Although this airport configuration only occurs about 20% of the time, studies have shown that it accounts for roughly half the delays at LGA [64].

5.4.2 LaGuardia Evaluation Method

The data was divided similar to the way we divided previous data using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period.

The LaGuardia Airport demand chart in Figure 5.9 shows peaks at 0700, 1000, 1400, and 1800 local time, so they were chosen as the points to analyze. The most common AAR for LaGuardia is 40 followed by 38 and 37. We were able to find a function that divides 40 from the rest of the data, however we were unable to have the same results for 38 and 37.
Figure 5.8: LaGuardia Airport Map

[43]
When trying to separate 38 and 37 AAR data from the rest, the SVM produced a predictor vector $w$ with all zeros and an $b$ of -1, so the SVM was not effective in determining the differences between these AARs. Therefore we used the fourth most popular AAR, 35, to separate the rest of the data.

### 5.4.3 LaGuardia Results

The results of the training and testing data for LaGuardia are found in Table 5.17 and Table 5.18.

As discussed earlier, LaGuardia is a congested airport. Since the queue is full, any small disruption, weather or otherwise, can cause large arrival delays. Therefore weather alone may not adequately address the reason for the delay. Despite this, Table 5.16 shows the training data algorithm was correct 71% of the time. Table 5.17 shows that the testing data was not as successful with only 59% of the data correctly predicted, but considering the other capacity issues at LaGuardia, it is still encouraging.

Table 5.19 and Table 5.20 show how successful the tool will predict future AARs at
### Table 5.16: LaGuardia Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>36</td>
<td>0.32</td>
<td>0.89</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.91</td>
<td>0.55</td>
<td>0.82</td>
<td>0.73</td>
<td>0.80</td>
</tr>
<tr>
<td>1000</td>
<td>36</td>
<td>0.41</td>
<td>0.84</td>
<td>0.64</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.90</td>
<td>0.55</td>
<td>0.82</td>
<td>0.70</td>
<td>0.79</td>
</tr>
<tr>
<td>1400</td>
<td>36</td>
<td>0.35</td>
<td>0.88</td>
<td>0.67</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.93</td>
<td>0.35</td>
<td>0.77</td>
<td>0.68</td>
<td>0.76</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>0.27</td>
<td>0.92</td>
<td>0.67</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.98</td>
<td>0.11</td>
<td>0.70</td>
<td>0.73</td>
<td>0.70</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>0.71</td>
<td>0.71</td>
<td>0.75</td>
<td>0.67</td>
<td>0.71</td>
</tr>
</tbody>
</table>

### Table 5.17: LaGuardia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>36</td>
<td>0.05</td>
<td>0.97</td>
<td>0.43</td>
<td>0.69</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.07</td>
<td>0.97</td>
<td>0.80</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>1000</td>
<td>36</td>
<td>0.11</td>
<td>0.98</td>
<td>0.67</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.08</td>
<td>0.94</td>
<td>0.77</td>
<td>0.29</td>
<td>0.33</td>
</tr>
<tr>
<td>1400</td>
<td>36</td>
<td>0.17</td>
<td>0.98</td>
<td>0.75</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.32</td>
<td>0.83</td>
<td>0.87</td>
<td>0.27</td>
<td>0.44</td>
</tr>
<tr>
<td>1800</td>
<td>36</td>
<td>0.19</td>
<td>0.96</td>
<td>0.64</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.79</td>
<td>0.43</td>
<td>0.76</td>
<td>0.46</td>
<td>0.68</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>0.26</td>
<td>0.92</td>
<td>0.77</td>
<td>0.55</td>
<td>0.59</td>
</tr>
</tbody>
</table>
LaGuardia. If the algorithm indicated that the data point was less than 36, then the tool predicted an AAR of 32. An AAR of 32 was calculated based on the weighted average of all AARs under 36. If the data point is greater than or equal to 36 and less than 39, then the tool predicts an AAR of 37 which is the weighted average of all AARs between 36 and 39. In all other instances the tool predicted an AAR of 40.

Table 5.20 shows the delays for LaGuardia. The training data in Table 5.18 performs very well with the lower or upper values correctly predicted at least 65% of the time. The testing data in Table 5.19 was not very accurate for the higher value, but it still indicated, with good accuracy, whether or not the airport will experience irregular operations. Again, this loss of performance may be due to the other congestion factors that affect LaGuardia Airport.

### 5.5 John F. Kennedy International Airport

Previous sections of this paper examined two of the major airports that serve the New York metropolitan area. This section examines the third major airport, John F. Kennedy
Table 5.19: Tool Results for LaGuardia Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>32</td>
<td>43%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>29%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>33%</td>
<td>93%</td>
</tr>
<tr>
<td>1000</td>
<td>32</td>
<td>67%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>33%</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>30%</td>
<td>92%</td>
</tr>
<tr>
<td>1400</td>
<td>32</td>
<td>75%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>33%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>27%</td>
<td>71%</td>
</tr>
<tr>
<td>1800</td>
<td>32</td>
<td>64%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>49%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>46%</td>
<td>28%</td>
</tr>
</tbody>
</table>

Table 5.20: LaGuardia Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>0700</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>6</td>
</tr>
<tr>
<td>1000</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>29</td>
</tr>
<tr>
<td>1400</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>48</td>
</tr>
<tr>
<td>1800</td>
<td>Low</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>85</td>
</tr>
</tbody>
</table>
International Airport (JFK). JFK airport is the newest and busiest of New York City’s three primary commercial airports. It is the top international air passenger gateway to the United States. JFK’s outbound international travel accounted for 17% of all U.S. travelers who went overseas in 2004, the largest share of any U.S. airport [65].

5.5.1 Airport Layout

A seen in Figure 5.10 there are two pairs of parallel runways, four in all, surround the airport’s central terminal area: 4L-22R, 4R-22L, 13L-31R and 13R-31L. Runway 13R-31L is the second longest commercial runway in North America, at a length of 14,572 ft.

5.5.2 JFK Evaluation Method

The data was divided similar to the way we divided previous data using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period. Figure 5.11 shows that JFK has a different demand pattern then other
airports. The maximum AAR for JFK is 53, but it can be seen from the chart that this is only approached at 1600 local time. At all other times the arrival demand is low compared to arrival capacity. This is attributed to JFK being primarily an international airport and a large number of overseas flights arrive at around 1600 local time. This made it difficult for the SVM to work for any time period except 1600, since AARs are seldom reduced to a level close to the demand.

The most popular AAR at 0600 and 1900 is 35 and the AAR is at least 32 86% of the time. At times the AAR at 0600 is 53, but this only occurs 6% of the time. At 1200, 33 is the most popular AAR and the AAR is at least 32 84% of the time. This makes it difficult to find a separating hyperplane for any values outside of the range of 32 through 35. Therefore the algorithm can not distinguish between normal and irregular operations.

5.5.3 JFK Airport Results

The results of the training and testing data for 1600 local at JFK are found in Table 5.22 and Table 5.23.
Table 5.21: JFK Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>49</td>
<td>0.30</td>
<td>0.92</td>
<td>0.74</td>
<td>0.63</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>0.93</td>
<td>0.36</td>
<td>0.80</td>
<td>0.65</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 5.22: JFK Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>49</td>
<td>0.30</td>
<td>0.95</td>
<td>0.70</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>0.91</td>
<td>0.51</td>
<td>0.83</td>
<td>0.68</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Although the method is unable to make any predictions for any time period except 1600, the data from Table 5.21 and Table 5.22 show that the training data was correct 78% of the time and the test data 80% of the time.

If the algorithm indicated that the data point was less than 49, then the tool predicted an AAR of 35. An AAR of 35 was calculated based on the weighted average of all AARs under 49. If the data point is greater than or equal to 49 and less than 53, then the tool predicts an AAR of 51 which was the most popular AAR in the time period. In all other instances the tool predicted an AAR of 53 which was the second most popular in the time period.

Table 5.23 shows the usual success on the edges, with less accuracy at the middle value. Table 5.24 shows accuracy at all values which was very encouraging. The tool is only

Table 5.23: Tool Results for JFK Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>35</td>
<td>74%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>39%</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>66%</td>
<td>14%</td>
</tr>
</tbody>
</table>
Table 5.24: Tool Results for JFK Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>35</td>
<td>70%</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>59%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>67%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Table 5.25: JFK Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
<th>35</th>
<th>51</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>Low</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td>14</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>1000</td>
<td>Low</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td>20</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>1400</td>
<td>Low</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>17</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td>31</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>1800</td>
<td>Low</td>
<td></td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>23</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td>55</td>
<td>31</td>
<td>24</td>
</tr>
</tbody>
</table>
effective at 1600, but the demand at other time periods is very low and not subjected to reduction in the normal AAR due to weather or congestion. Table 5.25 is based on actual AARs and therefore can show the delays for all time periods and AARs.

5.6 Ronald Reagan National Airport

Reagan National Airport is a public airport located three miles south of the central business district of Washington, D.C. It is notable for being the nearest commercial airport to the District of Columbia. With a handful of exceptions, flights are restricted to destinations within 1,250 miles, in an effort to control aviation noise and to drive air traffic to the larger but more distant Washington Dulles International Airport. In 2006, the airport served approximately 18.5 million passengers [66].

Based on the data collected from the ASPM database, on April 28, 2002, the maximum AAR went from 24 to 40. This is probably because the Department of Transportation restored National to full service after closing the airport completely after the terrorist attacks of September 11th.

5.6.1 Airport Layout

A seen in Figure 5.12 there are three intersecting runways all oriented in a basically a north south direction. The runway layout, limited due to the location and orientation of the airport, has otherwise changed little since 1955 except for the 1956 closure of a fourth, east-west runway now used for taxiing and aircraft parking.

5.6.2 Reagan National Airport Evaluation Method

The data was divided similar to the way we divided previous data using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period.

The Reagan Airport demand chart in Figure 5.13 shows peaks at 0800, 1300, 1600, and 1800 local time, so they were chosen as the points to analyze. The most common AAR for
Figure 5.12: Reagan National Airport Map

[43]

Figure 5.13: DCA Hourly Demand
Table 5.26: Reagan National Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>31</td>
<td>0.33</td>
<td>0.97</td>
<td>0.67</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.37</td>
<td>0.94</td>
<td>0.71</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>1300</td>
<td>31</td>
<td>0.32</td>
<td>0.97</td>
<td>0.66</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.37</td>
<td>0.95</td>
<td>0.71</td>
<td>0.81</td>
<td>0.79</td>
</tr>
<tr>
<td>1600</td>
<td>31</td>
<td>0.32</td>
<td>0.98</td>
<td>0.69</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.32</td>
<td>0.94</td>
<td>0.67</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>1800</td>
<td>31</td>
<td>0.30</td>
<td>0.98</td>
<td>0.68</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.37</td>
<td>0.95</td>
<td>0.73</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.34</td>
<td>0.96</td>
<td>0.70</td>
<td>0.85</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Reagan is 36 followed by 40 and then 32. When trying to separate 40 from the rest of the data, the SVM produced a predictor vector $w$ with all zeros and an $b$ of -1, so the SVM was not effective in determining the differences between these AARs. The SVM effectively separates data at 36 and 32, so these AAR values were chosen as break points for the data. Since Reagan did not return to full operational capacity until Spring of 2002, no data before April 28, 2002 was included in the analysis. This reduced the training data set from 1826 points to only 1709.

5.6.3 Reagan National Results

The results of the training and testing data for Reagan National are found in Table 5.27 and Table 5.28.

Table 5.26 shows that the algorithm is 84% accurate when predicting the training data and Table 5.27 shows that the algorithm is 76% accurate when predicting the testing data.

For Reagan National, if the algorithm indicated that the data point was less than or equal to 31, then the tool predicted an AAR of 27. An AAR of 27 was calculated based on the weighted average of all AARs under 31. If the data point is greater than 31 and less than or equal to 35, then the tool predicts an AAR of 33 which is the weighted average of all AARs between 32 and 36. In all other instances the tool predicted an AAR of 38 which
Table 5.27: Reagan National Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>31</td>
<td>0.43</td>
<td>0.96</td>
<td>0.50</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.16</td>
<td>0.98</td>
<td>0.87</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td>1300</td>
<td>31</td>
<td>0.31</td>
<td>0.95</td>
<td>0.33</td>
<td>0.95</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.17</td>
<td>0.98</td>
<td>0.88</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td>1600</td>
<td>31</td>
<td>0.33</td>
<td>0.95</td>
<td>0.31</td>
<td>0.95</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.16</td>
<td>0.98</td>
<td>0.87</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td>1800</td>
<td>31</td>
<td>0.36</td>
<td>0.95</td>
<td>0.31</td>
<td>0.96</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.12</td>
<td>0.98</td>
<td>0.83</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.18</td>
<td>0.96</td>
<td>0.63</td>
<td>0.77</td>
<td>0.76</td>
</tr>
</tbody>
</table>

is a weighted average of all AARs greater than 35.

The training data in Table 5.28 performs very well with the lower or upper values correctly predicted at least 66% of the time. The testing data in Table 5.29 was not very accurate for the lower value, but it still indicated, with good accuracy, whether or not the airport will experience irregular operations. This poor performance for the testing data may be caused by the fact that Reagan National has fewer capacity reductions than other airports in the study, therefore capacity reduction is a rare event. Table 5.30 shows the delays for Reagan National.

### 5.7 Dulles International Airport

Washington Dulles International Airport is located 25 miles west of Washington. Although Reagan National Airport had been open only since 1941, the need for a second airport to serve the National Capital Area had become apparent shortly after the end of World War II. To meet the growing demand for airport capacity, Congress passed the second Washington Airport Act of 1950 to provide for the construction, protection, operation, and maintenance of a public airport in or in the vicinity of the District of Columbia [67].
Table 5.28: Tool Results for Reagan National Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>27</td>
<td>67%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>23%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>80%</td>
<td>86%</td>
</tr>
<tr>
<td>1300</td>
<td>27</td>
<td>66%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>20%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>81%</td>
<td>86%</td>
</tr>
<tr>
<td>1600</td>
<td>27</td>
<td>69%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>20%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>80%</td>
<td>87%</td>
</tr>
<tr>
<td>1800</td>
<td>27</td>
<td>68%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>30%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>80%</td>
<td>86%</td>
</tr>
</tbody>
</table>

Table 5.29: Tool Results for Reagan National Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>27</td>
<td>50%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>25%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>60%</td>
<td>91%</td>
</tr>
<tr>
<td>1300</td>
<td>27</td>
<td>33%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>75%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>59%</td>
<td>91%</td>
</tr>
<tr>
<td>1600</td>
<td>27</td>
<td>31%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>83%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>60%</td>
<td>90%</td>
</tr>
<tr>
<td>1800</td>
<td>27</td>
<td>31%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>100%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>59%</td>
<td>91%</td>
</tr>
</tbody>
</table>
Table 5.30: Reagan National Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>0800</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>13</td>
</tr>
<tr>
<td>1300</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>39</td>
</tr>
<tr>
<td>1600</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>39</td>
</tr>
<tr>
<td>1800</td>
<td>Low</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>26</td>
</tr>
</tbody>
</table>

5.7.1 Airport Layout

A seen in Figure 5.14 Dulles had two north-south parallel runways, each 11,500 feet long, 150 feet wide, and separated by 6,700 feet and a third northwest-southeast runway, 10,000 feet long and 150 feet wide. All runways had standard instrument landing systems (ILS) for landings, high-speed turnoffs to increase runway availability.

5.7.2 Dulles International Airport Evaluation Method

The data was divided similar to the way we divided previous data using the SVM. To determine these points, ASPM data was analyzed to find the most common AAR for the given time period.

The Reagan Airport demand chart in Figure 5.15 shows peaks at 0700, 1100, 1500, and 2000 local time, so they were chosen as the points to analyze. From April through August of 2004, a 10,000-foot-long, 150-foot-wide runway, built when the airport opened in 1962, had to be reconstructed because of wear and tear, airport officials said. The loss of one of the airport’s three runways during the heavy summer travel season contributed to delays at Dulles, but officials said the work had to be done during warm weather [68].
Figure 5.14: Dulles International Airport Map

Figure 5.15: IAD Hourly Demand
Table 5.31: Dulles Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>Under 80</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.15</td>
<td>0.96</td>
<td>0.63</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>1100</td>
<td>Under 80</td>
<td>0.23</td>
<td>0.97</td>
<td>0.80</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>1500</td>
<td>Under 80</td>
<td>0.27</td>
<td>0.94</td>
<td>0.62</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.19</td>
<td>0.93</td>
<td>0.60</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>2000</td>
<td>Under 80</td>
<td>0.23</td>
<td>0.94</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.21</td>
<td>0.93</td>
<td>0.62</td>
<td>0.69</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.20</td>
<td>0.96</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
</tbody>
</table>

of this construction, the AAR data had to be normalized to use all of the data during the training period. This produced as set of AARs that were not grouped as tightly as those airport where airport maintenance conditions were constant.

Figure 5.15 also shows large demand during the peak periods followed by a large decrease in activity. The most popular AAR at Dulles was 80, while the second most popular AAR was 90. This does not imply that Dulles had more delays, but based on the data described in Figure 5.15, average demand only approaches maximum capacity at 1500 local. Because of this the data was grouped in three bins; one that included all AARs under 80, one that included all AARs of 80, and a final bin with all AARs greater than 80.

5.7.3 Dulles International Results

The results of the training and testing data for Dulles are found in Table 5.32 and Table 5.33.

Both Table 5.31 and Table 5.32 seem to show that the algorithm is effective in predicting the future AAR. Upon further review, the sensitivity for the over 80 data at 1100 is 0 and the specificity is 1. This means that the algorithm did not produce a vector to divide the data. Therefore we can not make any conclusions to whether the data had an AAR over 80 at 1100. Further issues are discovered when reviewing the accuracy in Table 5.33 and
Table 5.32: Dulles Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Divider</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>Under 80</td>
<td>0.20</td>
<td>1.00</td>
<td>1.00</td>
<td>0.61</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.12</td>
<td>0.97</td>
<td>0.60</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>1100</td>
<td>Under 80</td>
<td>0.14</td>
<td>0.95</td>
<td>0.69</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>1500</td>
<td>Under 80</td>
<td>0.14</td>
<td>0.98</td>
<td>0.87</td>
<td>0.51</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.16</td>
<td>0.95</td>
<td>0.63</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>2000</td>
<td>Under 80</td>
<td>0.09</td>
<td>0.96</td>
<td>0.75</td>
<td>0.46</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Over 80</td>
<td>0.36</td>
<td>0.89</td>
<td>0.56</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.15</td>
<td>0.96</td>
<td>0.71</td>
<td>0.63</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 5.34.

For Dulles, if the algorithm indicated that the data point was less than 80, then the tool predicted an AAR of 68. An AAR of 68 was calculated based on the weighted average of all AARs under 80. If the data point is greater than 80 then the tool predicts an AAR of 90 which is the weighted average of all AARs over 80. In all other instances the tool predicted an AAR of 80. Table 5.35 summarizes the delays for Dulles.

The results for Dulles are not very encouraging. The most probable AAR that should occur is 80, however the accuracy rate for all data types, times, and AAR prediction are very low. Earlier analysis showed that middle values were the most difficult to predict, therefore having the most probable value as a middle value is a problem for the linear SVM algorithm.
Table 5.33: Tool Results for Dulles Training Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>68</td>
<td>71%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>43%</td>
<td>76%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>63%</td>
<td>7%</td>
</tr>
<tr>
<td>1100</td>
<td>68</td>
<td>80%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>40%</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1500</td>
<td>68</td>
<td>62%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>41%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>61%</td>
<td>11%</td>
</tr>
<tr>
<td>2000</td>
<td>68</td>
<td>67%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>35%</td>
<td>76%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>62%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Table 5.34: Tool Results for Dulles Testing Data

<table>
<thead>
<tr>
<th>Time</th>
<th>AAR</th>
<th>Percent of Predictions Correct</th>
<th>Percent of Actual AAR Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td>68</td>
<td>84%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>34%</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>60%</td>
<td>6%</td>
</tr>
<tr>
<td>1100</td>
<td>68</td>
<td>69%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>27%</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1500</td>
<td>68</td>
<td>87%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>16%</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>63%</td>
<td>9%</td>
</tr>
<tr>
<td>2000</td>
<td>68</td>
<td>75%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>19%</td>
<td>76%</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>56%</td>
<td>18%</td>
</tr>
</tbody>
</table>
Table 5.35: Dulles International Delay Predictions

<table>
<thead>
<tr>
<th>Time</th>
<th>Delay</th>
<th>Predicted AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>68</td>
<td>80</td>
</tr>
<tr>
<td>0800</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>4</td>
</tr>
<tr>
<td>1300</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>24</td>
</tr>
<tr>
<td>1600</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>20</td>
</tr>
<tr>
<td>1800</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>48</td>
</tr>
</tbody>
</table>
Chapter 6: Strategies for the Delay Tool

The Terminal Aerodrome Forecast can be used to create a tool to predict Aircraft Arrival Rates, Ground Delay Programs, and delay. This chapter will show how the tool can be used by airlines to improve operations and how the Air Traffic Control System Command Center (ATCSCC) can modify the Military Decision Making Process (MDMP) to create a plan for the operational day. The first part of this chapter focuses on airline that runs a hub and spoke system like American Airlines while the second part focuses on an airline that runs a skip-stop system such as Southwest Airlines. The final part of the chapter will show how the MDMP can be modified to be used in civilian air traffic management and how the Weather Delay Prediction Tool is integrated into that planning.

6.1 Hub and Spoke Operations

Hub and spoke networks have become the most popular type of airline scheduling. In this type of scheduling, several points of departure feed into a single hub airport from which connecting flights carry passengers to their final destination. The advantage of the cross-connections is the multiplier effect as to the number of city pair that can be served. However, airports that are designated as the “hub” are subjected to increased congestion that are exasperated by irregular operations [6].

Traffic at airline connecting hubs usually operates in four or more distinct daily peaks, lasting an hour or less a piece. During these peaks, the airport briefly operates near its physical runway capacity. However, the airports terminal and runways are under-utilized for the majority of the day. If inclement weather reduces runway capacity or closes the airport during a flight bank, serious delay results for the airline and its passengers.

If airport arrival capacity is reduced because of inclement weather, flights are usually
delayed on the ground through a ground delay program. The passengers on the delayed flights must wait for the weather at the hub to clear. If the original connecting flight was able to depart the airport near its scheduled time, the passenger may have missed his connection; he may have to wait for the next flight to his final destination, causing even more delay.

If inclement weather affects a departing flight bank at large hub airports, extreme departure delays with large queues can result. Bad weather at a connecting hub creates ripples or a cascading effect throughout an airline’s entire route system. A flight delayed an hour at a hub is an hour late arriving at the next spoke city. Subsequently, this flight would depart from the spoke city, heading to another hub still an hour late. Each passenger served by this aircraft, for the rest of the day, would be an hour late [12].

So how can an airline avoid delays caused by irregular operations at one of their hub cities? One solution is to use the TAF Delay Prediction Tool and couple it with the concept of a reliever hub.

6.1.1 Concept of the Reliever Hub

Meyer et al. (1998) [12] introduces the reliever and floating hub concept in a 1998 paper. Since most airline schedules are made without regard to unexpected daily changes due to severe weather conditions, there is very little slack time which means that any delay early in the day is likely to affect the schedule for the rest of the day unless the airline can take effective steps to correct the problem. Most carriers have developed procedures to follow in the event of unexpected disruptions in operations. However, most of these procedures are implemented manually, with little or no reliance on automated decision support systems [6].

The reliever hub is a strategy to reassign and optimize airport and airline schedules when experiencing a disruptive disturbance at a major hub airport and still maintain reasonable service. There are two options within the reliever hub concept. Figure 6.1 shows an example of a hub and spoke system. Option one is to temporarily use a nearby airport
Figure 6.1: Hub and Spoke System
to act as a connecting hub, which can reduce the delays caused by a capacity reduction at the major hub. Figure 6.2 shows city 1 as a reliever hub. All cities to the west of the hub are sent to the reliever hub while all cities to the east continue to go to the main hub which reduces the demand on the main hub and decreases delays within the system. Service from city 1 to the hub would probable have to use a larger plane or more flights to insure passengers that need to get to the eastern cities or to the hub city arrive at their final destination. Option two would be over-flying the of the hub and swapping larger aircraft onto other routings throughout the system enabling passengers to connect through alternative hubs [12]. Figure 6.3 shows cities 2, 5, 6, and 9, where the planes fly over the hub and go directly to another city. Routes that skip the hub must be chosen carefully to ensure that the fewest amount of passengers are impacted. Also the flight from cities 3, 4, 7, and 8 to the hub city will probably need to have a larger plane, so those customers headed to the
hub city or other cities are quickly routed to their final destination.

6.1.2 Integrating the TAF Delay Tool into the Reliever Hub Concept

Determining the best location for a reliever hub is a complex task. Even if the alternate hub is chosen ahead of time, the airline still must reconfigure schedule and passenger itinerary to minimize the total delay. Whether this reconfiguration is done by hand or is automated, it still requires time and employee manpower. It also requires a decision to use this manpower. Manpower has a cost associated with it, so the airline manager has to reasonably sure that the labor cost will help reduce any future loss due to extended delays. Because the TAF Delay Prediction Tool predict future delays it provides time and justification. Since the tool has been trained by historical data, it makes an AAR prediction based on what happened with similar TAFs in the past. Now the manager has justification to begin planning for the
Because the manager can enter the TAF a day in advance, there is now time to implement the plan by rescheduling flights and even informing passengers of any changes. Larger or smaller aircraft can be swapped to account for the change in passenger. This time also allows ground crews at the reliever hub more time to ready themselves for unusually high activity. The reliever hub may not have the permanent infrastructure to support an increase in passengers, therefore temporary solutions may have to be implemented. Since there may not be enough gates, planes may have to be serviced on the apron. This may require the use of buses and bus drivers to drive passengers from the planes to the terminal. Temporary shelter may have to be set up to shield passengers from the heat or cold.

### 6.1.3 Example Planning Day

As part of the airline standard operating procedure, the airline operations manager runs the TAF Delay Prediction Tool and finds that a major hub within the airline system will experience a capacity reduction between 1500 and 2400 local time. The tool predicts that the AAR will go from 50 to 35 during that period which equates to a 30% reduction in operations at that airport. This means that 15 fewer planes per hour will be able to land at the airport. This is a major hub and the airline accounts for 80% of the traffic in and out of the airport. Therefore the manager concludes that about 12 of the company’s aircraft will be affected each hour. The manager decides to activate a reliever hub to reduce the demand at the hub.

The object of the reliever hub is not to shut the other main hub down completely, but instead reduce the demand which will reduce the delay. The manager estimates that the reliever hub is going to have to take on an additional 12 flights per hour. From this estimate gate capacities are determined. If this rate exceeds the permanent gate capacity, then alternate gates are going to have to be created. These alternate gates will use space on the apron to unload the passengers and luggage. Buses or vans will then be required to transport the passengers to the gates. If the weather is inclement or it is extremely...
hot or extremely cold then the passengers should be kept on the plane until buses arrive or temporary shelters should be set up to keep passengers out of the elements. Some of the passengers at the reliever hub, had the main hub as their final destination. Typically a reliever hub should be an airport within the airlines network of cities that it serves, therefore there should be a flight available to take the passengers from the reliever hub to the main hub. Since the reliever hub reduced the demand at the main, then hopefully flight from the reliever hub to the main hub will experience minimal delay. A plane swap to increase the size of the aircraft on the main hub flight may have to occur to account for the additional number of passengers.

Setting up a reliever hub can take up time and resources. If American Airlines were to use Austin as reliever hub for Dallas-Fort Worth, it would cost $9.7 million annually [12]. Because of this it is important to use a tool that will identify the potential need and increase the available time to react. The potential benefits are huge, because although the Austin reliever hub required an significant initial outlay, savings were estimated to be $14.2 million dollars annually [12].

6.2 Skip-Stop Operations

Skip-stop operations are when an airline schedules provide service to a number of stations $A_1, A_2, \ldots, A_n$, by scheduling flights $A_1 \rightarrow A_3 \rightarrow A_5 \rightarrow A_n$ and $A_2 \rightarrow A_4 \rightarrow A_6 \rightarrow A_{n-1}$ (as a hypothetical example). In this type of schedule one or more of the intermediate stops are “skipped” and service to those stations is provided by another flight. This way fast service can be provided between intermediate stations ($A_1$ and $A_3$ for example). On the other hand, no service is provided between consecutive cities ($A_1$ and $A_2$, for example) [6].

6.2.1 Southwest Airlines Operations

One example of an airline that runs a skip-stop schedule is Southwest Airlines. Southwest Airlines is a low cost airline based in Dallas that has carried more customers than any
other U.S. airline since August 2006 for combined domestic and international passengers according to the U.S. Department of Transportation’s Bureau of Transportation Statistics [69]. Southwest Airlines is one of the world’s most profitable airlines and in January 2007, posted a profit for the 34th consecutive year [70]. Southwest does not use the more traditional hub and spoke flight routing system of most other major airlines, preferring instead the “Point to Point” or what was previously described as a skip-stop system. An average of 80% of Southwest passengers are local passengers, meaning only 20 percent of all passengers are connecting passengers. This is significantly higher than most airlines, where passengers often connect in hub cities [71].

An airline like Southwest would have difficulty using a system that utilizes a reliever hub. Reliever hubs are only effective when most of the passengers intend to connect. If Southwest tried this, and sent a plane to alternate destination, then 80% of the passengers would be in the wrong place. Even if it was possible to service all of the connecting passengers at the alternate airport, the airline will further delay the 80% that want to travel to the original destination airport. This alternate hub technique would not reduce delay in the Southwest system, so some other delay reduction strategy must be developed.

### 6.2.2 Float Plane Operations

On October 17, 2006, the ATCSCC issued a GDP at Philadelphia to begin at 1300 local time. For that same day the Weather Delay Prediction Tool predicted a AAR of 36, the lowest predicted value possible, for the same time period as the GDP. To highlight the effects on an airline that runs skip-stop operations, all Southwest flights that at some point stop in Philadelphia were plotted on a spreadsheet in the form of a flow sheet seen in Appendix H. An examination of Appendix H shows that at about 1300, flights arriving at Philadelphia started to experience delay. Because of the nature of the Southwest schedule, these delays were passed on to flights further in the day that included cities that seemed to the average passenger, to have no connection to Philadelphia. For instance tail number N618WN was one hour and fifty-five minutes late on a flight from Providence to Baltimore-Washington.
This had nothing to do with weather in either of the two cities, but instead was caused by delay of a previous flight from Jacksonville to Philadelphia. Due to the fact that most passengers on Southwest flights are local, using alternate hubs or skipping stops is not an option, therefore a different solution needs to be developed.

One option for reducing total delay is to use a “float plane”. A float airplane is prepositioned at an airport and is used to replace an airplane on a flight that is delayed and could not meet the scheduled departure time for the next flight. A float plane should be prepositioned at an airport where it is likely that the capacity, based on aircraft arrival rate (AAR), is reduced due to weather. At most airports the maximum arrival capacity is usually less than the maximum departure capacity [72]. Also, if a GDP is instituted at an airport, then the number of inbound aircraft is reduced. Therefore there are fewer planes on the ground needing to depart after a turn around, so it is easier for a plane to depart in poor weather then arrive. The float plane will take advantage of this and have a better chance to depart on time. The airplane that was replaced by the float plane becomes the new float plane and will replace another delayed aircraft in order to reduce delays. The float plane concept will also reduce delays further down the schedule since it will arrive at the next destination on time and therefore have a good chance to take off on time for the next flight.

6.2.3 Example Float Plane Scenario

To see how much delay can be reduced by using the float plane, refer to the October 17, 2006 flow sheet in Appendix H. The Weather Delay Prediction Tool predicts a reduced AAR at Philadelphia. Based on this prediction airline management decides that there is a good chance that there will be a GDP at Philadelphia, so airline operations send a float plane to Philadelphia. This float is used to replace the Southwest flight, tail number N362, headed for Fort Lauderdale scheduled to leave at 1615. With an on-time departure, this float will save 42 minutes in delay for the Fort Lauderdale flight and, assuming another on-time departure, save 22 minutes on a following flight to Baltimore-Washington International (BWI). Now
tail number N362 is the float plane and is available at 1648. At 1710 tail number N362 takes the flight to Pittsburgh from tail number N218. Assuming on-time departure, this substitution saves 161 minutes and the following flight to Orlando saves 60 minutes. This float plane concept has saved a total of 285 minutes in delay. If we assume there were 100 people on each of the four flights, then this technique has saved 28,500 passenger-minutes.

6.3 Air Traffic Control System Command Center Planning

The Weather Delay Prediction Tool provides a powerful resource to predict future airport capacities due to inclement weather. The information provided by this tool is useless though, unless it is integrated into a systematic planning process. As described earlier, this planning process can be modeled after the Military Decision Making Process (MDMP). The MDMP has four phases;

- Mission Analysis,
- Course of Action Development,
- Course of Action Analysis and Comparison, and
- Decision and Execution [31].

Here the MDMP is married to a civilian planning process for air traffic management. Also the Weather Prediction Tool and other planning tools are integrated into the planning process in order to produce at daily plan for the NAS.

6.3.1 Mission Analysis

Mission Analysis is the first step in the MDMP process which consists of the following actions;

- Gathering current status and conditions,
- Making assumptions as a substitute for facts if information is unknown,
• Analyze higher mission and intent which determines what tasks are required to accomplish the mission, and

• Issue commander’s guidance to focus subsequent staff planning [31].

For the military this means, that subordinate organizations receive a mission for higher, gather any additional information, and produce a mission statement for lower subordinate organizations. For air traffic management, the mission does not change, so a mission statement is proposed, “the mission of the Air Traffic Control Command Center is to safely and efficiently direct the movement of all aircraft within the system.” Success is defined as all aircraft safely arriving at their final destinations with the minimum delay or disruptions. This is the part where the ATCSCC planning staff begins to collect information and use the tools to derive conclusions on any impacts on the NAS. The operations officer (OO) would use the FACET to determine the potential congestion within the system. The intelligence officer (IO) would use the Weather Prediction Delay Tool to determine any weather problems that would reduce airport capacity. The resource (RO) and personnel officers (PO) would collect any available information that may have impact on flow. All of this information is collected and consolidated in order to be used in the next phase.

6.3.2 Decision Point Development

For the military, a course of action (COA) is a possible plan to accomplish the mission. To develop COAs, the staff must focus on key information necessary to make decisions and assimilate the data in mission analysis. While developing COAs, the staff goes through the following steps;

• Analyze the enemy force,

• Develop a scheme,

• Determine command and control means and minimum required maneuver control measures, and
• Prepare COA statements [31].

The enemy of civilian air traffic is any obstacle that restricts the flow of air traffic. Typically these obstacles are caused by weather or schedule congestion. To develop a scheme, these obstacles are identified by the staff estimates. Staff estimates are the collection and the consolidation of all information collected in the mission analysis phase. These estimates will point out the obstacles to efficient flow which create a set of issues which could potentially become decision points. After the commander reviews all potential issues, the commander selects those issues that are to be labeled as decision points. Once the decision points have been identified then the commander working along with the staff should develop a set of branch plans to counter the effects. These branch plans should be written in the form, “at decision point A, the commander should either do B, C, or D.”

6.3.3 Plan Refinement, Distribution, and Execution

During military planning the staff typically come up with several courses of action which are evaluated as part of the planning process. This step can be added, but it is only needed if the ATCSCC has more than one maneuver plan to reroute traffic around potential bad weather or congestion. All options can then be “wargamed”, where a simulation is run with the staff acting as the humans in the loop. During this simulation a recorder should document issues with all potential scenarios so that pros and cons can be evaluated later. If a concept is deemed inadequate in terms of feasibility, then the staff must modify the concept or the commander must select a new concept [31].

Once the plan is analyzed and finalized it should be distributed to the Air Route Traffic Control Centers and the airlines. This ensures that all stakeholders within the NAS know what the decision points are, the potential branch plans, and any routing issues. Distribution is vital since it not only provides predictability, but also lets subordinate users of the NAS to formulate their own plans for the day based on the national strategy.

The execution of the plan is in the hands of the operations officer (OO). The OO monitors the days operations, brings the commander into process at decision points, and
updates any unforeseen changes to the NAS status. The other members of the planning staff also continue to update their own estimates and begin planning for the next day. The staff is also on call to begin a quick planning session if any major unforeseen trouble occurs during the day.

6.3.4 Sample Planning Day

As early as 24 hours before the beginning of the NAS day, the planning staff begins by starting to prepare the estimates. The OO runs the FACET model and sees potential congestion at 2000 Zulu, also known as 4:00 PM local, over the Cleveland ARTCC. The OO records this as a potential issue. The IO runs the Weather Delay Prediction Tool and finds that delays are predicted at Philadelphia starting at 1900 Zulu time which is 2:00 PM local time. The RO and PO find no major issues with personnel or resources.

After the initial staff estimates are completed and approved by the Chief of Staff, they are presented to the commander. The commander reviews the issues form the estimates and designates the congestion at 2000 Zulu over Cleveland ARTCC and the weather at Philadelphia at 1900 Zulu as decision points. The commander has the option of developing branch plans to counter these issues, but instead defers to the staff to come up with solutions that will be reviewed later.

In consultation with the PO, the OO discovers that Cleveland ARTCC can increase manning for that time period to handle any congestion. If the increase manning does not alleviate the problem, then a branch plan to include miles-in-trail (MIT) restrictions is proposed. Miles-in-Trail restrictions fix the distance between each aircraft and the one in front of it on a specific jetway, in order to regulate the flow evenly as it enters a restricted region of airspace, improving the ability of controllers to manage the aircraft as they pass through bad weather.

For Philadelphia, it is determined that the predicted weather will reduce airport capacity be decreasing the aircraft arrival rate which will lead to arrival delays. To mitigate this possible disruption, the IO creates four potential branch plans;
• MIT restrictions for Washington ARTCC

• Establish a Flow Constrained Area (FCA) for the Cleveland ARTCC boundary

• GDP at PHL, or

• Do nothing.

An FCA is like a GDP except instead of targeting an airport, it targets airspace.

These decision points and the associated branch plans are briefed to the commander who approves them. Once the plan is approved, it is consolidated and distributed to the ARTCCs and the airlines. Now these NAS stakeholders know what may happen during the upcoming day and they can begin to plan based on their own interests and requirements. Based on the decision points the airlines may use float planes or alternate hubs to reduce the potential delays. Meanwhile the Cleveland ARTCC may begin to reallocate personnel resources to deal with upcoming congestion in their sectors.
Chapter 7: Conclusions and Future Work

Over the years airline delays have worsen so much so that the problem is attracting Presidential attention. During the 2007 Thanksgiving and Christmas travel season, the President attempted to alleviate these delays by opening up military airspace to civilian transport[73]. Although this may have provided some relief, it does not address the problem. The principle bottlenecks of the air traffic control system are the major commercial airports, of which at least a dozen currently operate near or above their point of saturation under even moderately adverse weather conditions [1]. Due to their cost and the environmental and noise issues associated with construction, it is unlikely that any new airports will be built in the near feature. Therefore, to make the National Airspace System run more efficiently, techniques to more effectively use the limited airport capacity must be developed.

Air Traffic Management has always been a tactical exercise, with decisions being made to counter near term problems[9]. Since decisions are made quickly, limited time is available to plan out alternate options that may better alleviate flow problems. Extra time means nothing when there is no way to anticipate future problems, therefore predictive tools are required to provide advance notice of future air traffic delays. This research provided two elements needed to more efficiently use the limited airport capacity. First, it introduced the Military Decision Making Process (MDMP) to civilian air traffic management. This process was modified for civilian use in order to simplify and standardize the planning process in order to develop options to handle potential problems. Second, the research described how to use Support Vector Machines (SVM) to predict future airport capacity. The Terminal Aerodrome Forecast (TAF) was used as an independent variable in the SVM to predict Aircraft Arrival Rates (AAR) which depict airport capacity. The research also showed that the AAR can be derived to determine Ground Delay Program (GDP) program rate and duration and passenger delay.
7.1 Weather Delay Prediction

The paper shows the possibilities of a Weather Delay Prediction Tool and what it can do to help NAS stakeholders. The algorithm is capable of classifying weather forecasts into three sets, where each set represents a specific AAR. Typically the highest AAR represents the airport during normal operations, while the two lower values represent reduced capacity due to weather or other congestion issues. Analysis showed that the SVM was more effective at predicting the normal AAR and the lower reduced capacity AAR. Therefore, for the weather delay prediction tool, it is appropriate to set a red, amber, or green scale to the output [74]. If the tool indicates green operations, then it is likely that the capacity at the airport will be at the maximum AAR and delays will be minimal. If the tool indicates red operations, then it is likely that the capacity at the airport will be significantly reduced and delays may be excessive. The Amber response indicates that the prediction is more uncertain, however, planners should prepared to have reduced operations at that airport. This appears to be the same rating system employed by the Weather Channel website, but it is also used by the military to rate progress of projects, describe the suitableness of terrain for armored vehicles, or any other situation that requires a general rating. Since the tool provides only a general assessment of airport capacity through AARs, then a general prediction of delays is also included based on AAR and time period. This tool helps the flying public know how long they can expect to be delayed due to weather.

Ground Delay Programs (GDP) will be predicted based on the tool prediction for each time period. The time periods for each airport were determined based on the demand peaks during the operational day. The two important pieces that came from a GDP are the programmed AAR and the duration. Programmed AAR is predicted based on the tool’s prediction for each time period. The length of the GDP is determined based on which time period are below the normal rate. For instance, the peak time periods at Philadelphia were 0800, 1200, 1600, and 1800. Therefore, if the AAR prediction at 1600 and 1800 were below normal and the AARs at 0800 and 1200, then we assume that the GDP begins half way between 1200 and 1600 at 1400 and lasts until the the end of the operational day at 2400.
7.2 Support Vector Machines

The analysis shows that the Support Vector Machine (SVM) is the best classification method to use to predict future AARs based on the Terminal Aerodrome Forecast (TAF). In all instances the receiver operating characteristic (ROC) curves show that the SVM is the right balance between sensitivity and specificity. SVMs also produce a simple linear prediction equation that can easily be updated when new historical information become available.

7.2.1 SVM Disadvantages

A disadvantage of the SVM is that it does not show if any factor has more influence on the outcome than another. For each individual prediction equation developed, there were same factors that were weighted higher than others. The prediction equation is not an intuitive answer. However, across all of the prediction equations, there was not a value that consistently had more influence than another. By the nature of the algorithm, recursive partitioning searches for the value that best divides the data, so if determining which factors have the most influence on the final solution, the recursive partitioning method is more appropriate. However, it can seen from Appendix I, that the rule set required to make prediction is much longer and will require longer programming time to update as the situation changes.

It is difficult to determine the effect of some of the data sets on the SVM. For instance, due to construction and airport upgrades at Atlanta and Dulles, some of the data was inconsistent. This analysis attempted to normalize the data to try to maintain a consistent data set. However there was no way to determine how this affected the actual results. Results from airports that required normalization were not as accurate as airports that had a consistent data set.

The SVM can not predict rare occurrences. For instance if a AAR rarely happened and the SVM tried to separate it from the rest of the data, a prediction vector of all zeros was the output with a \( b \) value of either 1 or -1. JFK International Airport had issues since it only consistently experienced the highest AAR at the 1600 time period. The high AAR
occurred occasionally at other time periods, but not often enough for the SVM to be able to separate it from the rest of the data.

7.2.2 Future Work on SVMs

Proper Data Set Size

One of the issues with the SVM is what is the right amount of training data needed to produce a prediction equation that produces the most accurate predictions without overfitting the data. The process in this paper was to develop a prediction equation and then compare it to the training data and then the testing data. The analysis used 57 factors in the prediction function. The number was based on the factors found in the TAF for four time periods. If more factors were added to the data, then the performance of the predictor function applied to the training data improved. Unfortunately this did not improve the performance of the testing data which indicates that the data was overfitted. This situation is similar to using a 8th order polynomial to predict a line with 8 variables. It performs well with the 8 variables, but has little prediction value for any new independent variable. An optimization algorithm could be developed to determine the proper mix of training data and factors. Additional data factors could be added by adding time periods and the size of the training data could be reduce, although not increased since TAF data only goes back as far as January 2002.

New York Area Airports

The SVM results for Newark Airport were encouraging, however the results were not as good at the other two major New York airports, LaGuardia and JFK. The runway configuration and congestion at LaGuardia and the odd AARs at JFK probably accounted for this. In the future it may make sense to analyze all three New York airports as one since they are relatively close together, share the same TRACON, and have similar weather. Since Newark offers the best results, historical analysis should be done to determine if there is a connection between reduced capacity at Newark and the capacity at the other two. If a
connection is made, then some simple regression technique may provide solid predictions at JFK and LaGuardia based on the Newark TAF.

Factors Other than Weather

This research focused on weather and using a forecast product to predict reduced airport capacity. Within the paper, it was discussed that other factor such as schedule congestion and runway construction also affect airport capacity. Further study should determine if there are any factors, besides weather factors, that can be added to the set of independent variables to produce a better predictive model. Techniques such as linear regression, have methods available to add or remove variables to the equation. At present though, there is no such standard process for SVMs other than adding the variable and testing the results to see if there is improvement. Developing this technique in itself would entail extensive analysis.

Data Mining Opportunities

Weather in itself does not cause delay. Delay is caused when controllers reduce the AAR into an airport in order to continue to safely land planes under proper separation standards. This is where the controller has the greatest impact on the process. Some controllers are more confident in their ability or have a greater tolerance for risk then others, therefore AARs are based on controller judgment. Using the SVM method, researchers can analyze controller actions based on a given forecast. Then, in consultation with aircraft manufactures and the airlines, researchers can determine if some of the capacity reductions were necessary. There is always a certain amount of risk associated with landing an airplane, so the intent is to see if controllers are allowing a reasonable amount of risk or are their restrictions too strict for today’s more advanced aircraft.
7.3 Air Traffic Management Planning

The Military Decision Making Process will provide a framework for planning operations within the National Airspace System. The only missing piece was a tool to help produce an estimate of the effects of weather on operations. Our results show that it is possible to create a tool that inputs the current TAF forecast and produces delay predictions hours in advance. There are two major barriers to air traffic flow, scheduled congestion and weather. Scheduled congestion can be predicted by using simulation tools that fly the current commercial airline schedules such as the Future ATM Concepts Evaluation Tool (FACET) developed by NASA. However these tools do not directly account for weather. The Weather Delay Prediction Tool helps with the weather part by providing the time and place of weather delays. Given this information, air traffic managers and airlines can create branch plans to counter any potential delays based on the historic performance of the system. This increased predictability should increase NAS efficiency.
Appendix A: Primal Formulation

param x {i in 1..1826, j in 1..57};  # observations
param y {i in 1..1826};   # Diagnosis

var w {j in 1..57} ;
var b {j in 1..57} ;
var sigh {i in 1..1826} >=0;

minimize total_var:
(1/2) * sum { j in 1..57 }
(w[j] * w[j]) + 1000 * sum { i in 1..1826 } sigh[i];

subject to cover {i in 1..1826} :
 y[i] * sum {j in 1..57} (w[j] * x[i,j] + b[j] ) + sigh[i] >=1;
for i=1:51
    for j=1:51
        SPTest(j,i)=exp(-(norm(SupportVector(i,:)-TestVectors(j,:)))^2/2) ;
    end
end
sum = alpha_y' * SPTest';
for k=1:51
    for l = 1:1826
        SPTestII(l,k)=exp(-(norm(SupportVector(k,:)-TestVectorsII(l,:)))^2/2) ;
    end
end
sumii = alpha_y' * SPTestII';
Appendix C: Dual Formulations

\begin{verbatim}
param G {i in 1..500, j in 1..500};  # observations
param y {i in 1..500};  # Diagnosis

var a {j in 1..500} >=0, <=10;

maximize total_var:
   sum { i in 1..500 }
      a[i] - (1/2) * sum { i in 1..500 } a[i] * sum { j in 1..500 } a[j] * G[j,i];

subject to two:
   sum { i in 1..500 } y[i] * a[i] = 0;
\end{verbatim}
Appendix D: Least Square Regression

\[
\text{param } x \{ i \text{ in } 1..1826, j \text{ in } 1..57 \}; \quad \# \text{ observations}
\]

\[
\text{param } y \{ i \text{ in } 1..1826 \}; \quad \# \text{ Diagnosis}
\]

\[
\text{var } w \{ j \text{ in } 1..57 \} ;
\]

\[
\text{var } b \ ;
\]

\[
\text{minimize total_var:}
\]

\[
\text{sum } \{ i \text{ in } 1..1826 \} \ (y[i] - \text{sum} \{ j \text{ in } 1..57 \} \ (x[i,j] * w[j]) - b)^2;
\]
Appendix E: Philadelphia Linear Regression Equations

\[ m = [-0.146, 1.508, -2.521, -6.137, 2.160, 0.000, 0.000, 1.043, -9.719, 0.002, \\
0.000, 0.001, 0.001, -0.219, 0.119, 0.487, 1.674, 0.000, -1.419, -4.432, \\
0.000, -1.061, 10.170, 0.002, 0.005, 0.005, 0.000, -0.297, -0.005, 0.077, \\
-0.965, 0.000, 0.348, -0.840, 4.891, -0.100, 0.000, 0.000, 0.001, -0.003, \\
0.004, -0.208, -0.087, 0.302, -0.491, -0.660, 0.445, 1.292, 0.000, 0.767, \\
0.000, -0.007, 0.002, 0.000, 0.002, 0.485, -0.461] \]

\[ b = 38.669 \]

\[ m = [-0.131, 0.701, -2.294, -0.578, 1.043, 0.000, -0.796, 0.203, -8.020, -0.002, \\
0.001, -0.002, 0.000, -0.604, 0.052, 0.537, -0.255, -4.921, 0.412, -2.815, \\
0.000, -0.977, 7.513, 0.003, 0.004, 0.004, 0.003, -0.235, 0.068, 0.062, \\
-1.691, -3.106, 2.183, 1.665, 2.029, -0.285, 0.000, -0.001, 0.003, 0.000, \\
0.003, 0.000, -0.145, 0.178, -0.034, 0.000, -0.985, 0.000, 0.000, 0.340, \\
0.000, -0.007, 0.000, -0.001, -0.002, 0.307, -0.390] \]

\[ b = 44.406 \]
\[ m = [-0.063, 0.141, -2.293, -2.235, 2.061, 0.000, 0.000, -0.691, 0.000, -0.004, \\
0.000, -0.003, 0.000, -0.482, 0.001, 0.571, -0.023, 0.000, -0.845, -2.525, \\
4.478, 0.609, 0.000, 0.004, 0.003, 0.006, 0.006, 0.503, 0.117, -0.215, \\
-1.778, -6.961, 1.339, 0.839, -4.912, -2.615, 0.000, -0.001, 0.001, -0.004, \\
-0.001, -0.749, -0.150, 1.058, 0.837, -1.249, -1.261, -0.803, 0.000, 1.020, \\
0.000, 0.000, 0.003, 0.004, 0.001, 0.105, -0.222] \]

\[ b = 42.925 \]

\[ m = [-0.054, 0.054, -2.452, -0.881, 1.707, 0.000, -1.442, -0.814, 6.045, -0.004, \\
0.000, -0.003, 0.000, -0.552, 0.004, 0.617, 0.938, -1.263, -0.695, -3.465, \\
3.900, 0.641, 0.000, 0.002, 0.004, 0.006, 0.006, 0.450, 0.112, -0.130, \\
-2.015, -5.320, 1.469, 1.072, -5.314, -2.887, 0.000, 0.000, -0.001, -0.004, \\
-0.001, -0.626, -0.156, 0.925, 0.000, -2.696, -0.689, 0.000, 0.000, 1.009, \\
0.000, 0.000, 0.003, 0.004, 0.003, 0.000, -0.290] \]

\[ b = 43.752 \]
Appendix F: Nearest Neighbor Distance Code

for i=1:1826
    for j=1:181
        SPTest(i,j)=norm(SupportVector(i,:)-TestVectors(j,:)) ;
    end
end
Appendix G: HDSS Access System

Station(s): sort by - Station(s) / State
- K01R - CLAIIBORNE RANG (AFS), LA (-)
- K01T - HIGH ISLAND, LA (-)
- K089 - GOAT ISLAND (RAMOS), ME (-)
- K0E4 - PAYSON, AZ (-)
- K0L3 - ZUMA BEACH, CA (-)
- K0W8 - CHINCOTEAGUE, VA (-)
- K0Y2 - STUNGEON BAY, WI (-)
- K128 - NEW CASTLES (CGLS), NH (-)
- K12N - AEROFLEX ANDOVER, NJ (-)

Enter Station ID below (overrides Station ids selected below):
(eg. KABR)

Bulletin Id(s): sort by - Bulletin Id(s) / Description
- ACUA6 - UNNUMBERED DEPRESSION + SUSPICIOUS AREA ADVISORY
- ACUS0 - Convective Outlook
- ACUS1 - SPC Mesoscale Discussion
- ACUS4 - TROPICAL WEATHER DISCUSSION
- ACUS5 - SEVERE STORM OUTLOOK NARRATIVE (AC)
- ACUS6 - TROPICAL WEATHER OUTLOOK AND SUMMARY
- ACUS9 - STORM SUMMARY
- ACUS6 - TROPICAL WEATHER SUMMARY
- ACUS7 - POST STORM HURRICANE REPORT
- ACZM7 - TEXT PRODUCT FOR INDEPENDENT SAMOA

Enter Bulletin ID(s) below (overrides bulletin ids selected above):
(eg. saus, acus7, wsus (enter 4 or 5 character bulletin ids, comma delimited))

Note: The bulletins in this system are in a variety of formats. We suggest that you not select a bulletin unless you are familiar with its format and usage.

Select MAX 1 Month Below

Start Date/Time:
(YYYY/MM/DD HH)
- To -
End Date/Time:
(YYYY/MM/DD HH)

Or -
Start Date:
(YYYY/MM/DD HH)
- To -
End Date:
(YYYY/MM/DD HH)

Output: FTP

Email Address: 
View Inventory?:  Yes

Continue With Selections  Reset Form
Appendix H: Southwest Airlines Flow Chart

This Appendix shows the Southwest Airlines flow chart for all planes that made a stop in Philadelphia. On the first page, the tail numbers of all of the planes are listed at the extreme left. On the top of the chart is the time of day (CDT).
<table>
<thead>
<tr>
<th>Time</th>
<th>0:00</th>
<th>1:30</th>
<th>3:00</th>
<th>4:30</th>
<th>6:00</th>
<th>7:30</th>
<th>9:00</th>
<th>10:30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- [Image]
<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
</tr>
<tr>
<td>11:00</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
</tr>
<tr>
<td>12:00</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
<td>CRP</td>
</tr>
</tbody>
</table>

**Note:**
- CRP: Classroom Research Project
- Schedule for the week starting from Monday and ending on Friday.
<table>
<thead>
<tr>
<th></th>
<th>12:00</th>
<th>12:30</th>
<th>13:00</th>
<th>13:30</th>
<th>14:00</th>
<th>14:30</th>
<th>15:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

151
<table>
<thead>
<tr>
<th>Time</th>
<th>15:30</th>
<th>16:00</th>
<th>16:30</th>
<th>17:00</th>
<th>17:30</th>
<th>18:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:30</td>
<td>API</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
</tr>
<tr>
<td>15:30</td>
<td>API</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
</tr>
<tr>
<td>15:30</td>
<td>API</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
<td>15:30</td>
</tr>
</tbody>
</table>

---

152
Appendix I: Recursive Partitioning Results

I.1 AAR 48 at 0800

Rule 1
Predicted class is '1'
IF Visibility (mi) ≤ 3.5 AND Wind Speed_1 ≤ 6.5 AND Visibility (mi)_1 ≤ 4.5 AND Scattered Ceiling_1 ≤ 27.5 THEN predicted = '1'

Results on training data
- Covered examples: 15
  - Class Estimated probability
    - -1 0.222
    - 1 0.778

Results on test data
- Covered examples: 4
  - Class Relative frequency
    - -1 1 0.500
    - 1 1 0.500

Rule 2
Predicted class is '1'
IF Wind Speed ≥ 4.5 AND Visibility (mi) ≤ 3.5 AND Wind Speed_1 ≥ 6.5 AND Visibility (mi)_1 > 4.5 AND Scattered Ceiling_1 ≤ 27.5 THEN predicted = '1'

Results on training data
- Covered examples: 18
  - Class Estimated probability
    - -1 0.250
    - 1 0.750

Results on test data
- Covered examples: 10
  - Class Relative frequency
    - -1 1 0.400
    - 1 1 0.600

Rule 3
Predicted class is '-1'
IF Wind Speed ≥ 4.5 AND Visibility (mi) ≤ 3.5 AND Wind Speed_1 ≤ 6.5 AND Visibility (mi)_1 > 4.5 AND Scattered Ceiling_1 ≤ 27.5 THEN predicted = '-1'

Results on training data
- Covered examples: 13
  - Class Estimated probability
    - -1 0.500
    - 1 0.500

Results on test data
- Covered examples: 3
  - Class Relative frequency
    - -1 1 0.000
    - 1 1 1.000

Rule 4
Predicted class is '1'
IF Visibility (mi) ≤ 3.5 AND Overcast Ceiling ≤ 11 AND Wind Speed_1 > 6.5 AND Scattered Ceiling_1 ≤ 27.5 AND Showers (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data
- Covered examples: 101
  - Class Estimated probability
    - -1 0.385
    - 1 0.615

Results on test data
- Covered examples: 42
  - Class Relative frequency
    - -1 1 0.357
    - 1 1 0.643

Rule 5
Predicted class is '-1'
IF Visibility (mi) ≤ 3.5 AND Overcast Ceiling ≤ 11 AND Wind Speed_1 > 6.5 AND Scattered Ceiling_1 ≤ 27.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 29
  - Class Estimated probability
    - -1 0.524
    - 1 0.476

Results on test data
- Covered examples: 7
  - Class Relative frequency
    - -1 1 0.571
    - 1 1 0.429

Rule 6
Predicted class is '1'
IF Visibility (mi) ≤ 3.5 AND Overcast Ceiling > 11 AND Wind Speed_1 > 6.5 AND Visibility (mi)_1 ≤ 5.5 AND Scattered Ceiling_1 ≤ 27.5 THEN predicted = '1'

Results on training data
- Covered examples: 17
  - Class Estimated probability
    - -1 0.800
    - 1 0.200

Results on test data
- Covered examples: 5
  - Class Relative frequency
    - -1 1 0.400
    - 1 1 0.600
Rule 7
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Overcast Ceiling > 11 AND Visibility (mi),_0 \leq 3.5 AND Wind Speed,\_1 > 6.5 AND Visibility (mi) \_1 > 5.5 AND Scattered Ceiling,\_1 \leq 27.5 THEN predicted = '1'

Results on training data
Covered examples: 8
Class Estimated probability
-1 0.200
1 0.800
Results on test data
Covered examples: 2
Class Relative frequency
-1 0.500
1 0.500

Rule 8
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Overcast Ceiling > 11 AND Visibility (mi),_0 > 3.5 AND Wind Speed,\_1 > 6.5 AND Visibility (mi) \_1 > 5.5 AND Scattered Ceiling,\_1 \leq 27.5 THEN predicted = '1'

Results on training data
Covered examples: 10
Class Estimated probability
-1 0.571
1 0.429
Results on test data
Covered examples: 6
Class Relative frequency
-1 0.500
1 0.500

Rule 9
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Broken Ceiling \leq 2 AND Scattered Ceiling,\_0 \leq 37.5 AND Scattered Ceiling,\_1 \geq 27.5 AND Thunder storms (Y/N),_2 \leq 0.5 AND Year \leq 3.5 THEN predicted = '1'

Results on training data
Covered examples: 13
Class Estimated probability
-1 0.600
1 0.400
Results on test data
Covered examples: 8
Class Relative frequency
-1 0.750
1 0.250

Rule 10
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Broken Ceiling \leq 2 AND Scattered Ceiling,\_0 \geq 37.5 AND Scattered Ceiling,\_1 \geq 27.5 AND Thunder storms (Y/N),_2 \leq 0.5 AND Year \leq 3.5 THEN predicted = '1'

Results on training data
Covered examples: 13
Class Estimated probability
-1 0.900
1 0.100
Results on test data
Covered examples: 6
Class Relative frequency
-1 0.833
1 0.167

Rule 11
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Broken Ceiling > 2 AND Scattered Ceiling,\_1 > 27.5 AND Thunder storms (Y/N),_2 \leq 0.5 AND Year \leq 3.5 THEN predicted = '1'

Results on training data
Covered examples: 18
Class Estimated probability
-1 0.600
1 0.400
Results on test data
Covered examples: 5
Class Relative frequency
-1 0.600
1 0.400

Rule 12
Predicted class is '1'
IF Visibility (mi) \leq 3.5 AND Scattered Ceiling,\_1 \geq 27.5 AND Thunder storms (Y/N),_2 \leq 0.5 AND Year > 3.5 THEN predicted = '1'

Results on training data
Covered examples: 38
Class Estimated probability
-1 0.667
1 0.333
Results on test data
Covered examples: 17
Class Relative frequency
-1 0.529
1 0.471
Rule 13
Predicted class is ‘-1’

IF Visibility (mi) \leq 3.5 AND Scattered Ceiling_1 > 27.5 AND Thunder storms (Y/N)\_2 > 0.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 12
  - Class: Estimated probability
  - 1: 0.500
  - 1: 0.500

Results on test data
- Covered examples: 7
  - Class: Relative frequency
  - 1: 0.857
  - 1: 0.143

Rule 14
Predicted class is ‘-1’

IF Visibility (mi) > 3.5 AND Visibility (mi) \leq 5.5 AND Cross Winds_0 \leq 0.5 AND Wind Speed_2 \leq 8.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 64
  - Class: Estimated probability
  - 1: 0.774
  - 1: 0.226

Results on test data
- Covered examples: 22
  - Class: Relative frequency
  - 1: 0.955
  - 1: 0.045

Rule 15
Predicted class is ‘-1’

IF Wind Speed \leq 5.5 AND Visibility (mi) > 3.5 AND Visibility (mi) \leq 5.5 AND Cross Winds_0 \leq 0.5 AND Wind Speed_2 > 8.5 AND Scattered Ceiling_2 \leq 47.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 9
  - Class: Estimated probability
  - 1: 0.800
  - 1: 0.200

Results on test data
- Covered examples: 9
  - Class: Relative frequency
  - 1: 0.889
  - 1: 0.111

Rule 16
Predicted class is ‘-1’

IF Wind Speed > 5.5 AND Visibility (mi) > 3.5 AND Visibility (mi) \leq 5.5 AND Cross Winds_0 \leq 0.5 AND Wind Speed_2 > 8.5 AND Scattered Ceiling_2 > 47.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 33
  - Class: Estimated probability
  - 1: 0.786
  - 1: 0.214

Results on test data
- Covered examples: 11
  - Class: Relative frequency
  - 1: 0.727
  - 1: 0.273

Rule 17
Predicted class is ‘-1’

IF Visibility (mi) > 3.5 AND Visibility (mi) \leq 5.5 AND Cross Winds_0 \leq 0.5 AND Wind Speed_2 \geq 8.5 AND Scattered Ceiling_2 > 47.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 13
  - Class: Estimated probability
  - 1: 0.857
  - 1: 0.143

Results on test data
- Covered examples: 4
  - Class: Relative frequency
  - 1: 1.000
  - 1: 0.000

Rule 18
Predicted class is ‘-1’

IF Visibility (mi) > 3.5 AND Visibility (mi) \leq 5.5 AND Showers (Y/N)\_0 > 0.5 AND Cross Winds_0 \leq 0.5 AND Wind Speed_2 \leq 5.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 17
  - Class: Estimated probability
  - 1: 0.600
  - 1: 0.400

Results on test data
- Covered examples: 9
  - Class: Relative frequency
  - 1: 0.889
  - 1: 0.111
Rule 19
Predicted class is "-1"
IF Visibility (mi) > 3.5 AND Visibility (mi) ≤ 5.5 AND Showers (Y/N) ≤ 0.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 7.5 AND Wind Speed_2 > 5.5 AND Overcast Ceiling_2 ≤ 95 AND Broken Ceiling_2 ≤ 50 THEN predicted = "-1"

**Results on training data**
- Covered examples: 11
  - Class Estimated probability
    - -1: 0.857
    - 1: 0.143

**Results on test data**
- Covered examples: 8
  - Class Estimated probability
    - -1: 0.875
    - 1: 0.125

Rule 20
Predicted class is "-1"
IF Visibility (mi) > 3.5 AND Visibility (mi) ≤ 5.5 AND Showers (Y/N) ≤ 0.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 > 7.5 AND Wind Speed_2 > 5.5 AND Overcast Ceiling_2 ≤ 95 AND Broken Ceiling_2 ≤ 50 THEN predicted = "-1"

**Results on training data**
- Covered examples: 38
  - Class Estimated probability
    - -1: 0.966
    - 1: 0.204

**Results on test data**
- Covered examples: 9
  - Class Estimated probability
    - -1: 0.955
    - 1: 0.444

Rule 21
Predicted class is "-1"
IF Visibility (mi) > 3.5 AND Visibility (mi) ≤ 5.5 AND Showers (Y/N) ≤ 0.5 AND Cross Winds_0 > 0.5 AND Wind Speed_2 > 5.5 AND Overcast Ceiling_2 > 95 AND Broken Ceiling_2 ≤ 50 THEN predicted = "-1"

**Results on training data**
- Covered examples: 22
  - Class Estimated probability
    - -1: 0.845
    - 1: 0.154

**Results on test data**
- Covered examples: 9
  - Class Estimated probability
    - -1: 0.667
    - 1: 0.333

Rule 22
Predicted class is "1"
IF Visibility (mi) > 3.5 AND Visibility (mi) ≤ 5.5 AND Showers (Y/N) ≤ 0.5 AND Cross Winds_0 > 0.5 AND Wind Speed_2 > 5.5 AND Overcast Ceiling_2 > 95 THEN predicted = "1"

**Results on training data**
- Covered examples: 8
  - Class Estimated probability
    - -1: 0.400
    - 1: 0.600

**Results on test data**
- Covered examples: 0
  - Class Estimated probability
    - -1: 0.000
    - 1: 0.000

Rule 23
Predicted class is "-1"
IF Visibility (mi) > 3.5 AND Visibility (mi) ≤ 5.5 AND Showers (Y/N) > 0.5 AND Cross Winds_0 > 0.5 THEN predicted = "-1"

**Results on training data**
- Covered examples: 11
  - Class Estimated probability
    - -1: 0.625
    - 1: 0.375

**Results on test data**
- Covered examples: 7
  - Class Estimated probability
    - -1: 0.571
    - 1: 0.429

Rule 24
Predicted class is "-1"
IF Visibility (mi) > 5.5 AND Broken Ceiling ≤ 27.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = "-1"

**Results on training data**
- Covered examples: 464
  - Class Estimated probability
    - -1: 0.970
    - 1: 0.030

**Results on test data**
- Covered examples: 200
  - Class Estimated probability
    - -1: 0.950
    - 1: 0.050
Rule 25
Predicted class is '+'1'

IF Visibility (mi) > 5.5 AND Broken Ceiling > 27.5 AND Cross Winds_0 ≤ 0.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = '-'1'

Results on training data

Covered examples: 86
  Class Estimated probability
    1  0.664
    0  0.336

Results on test data

Covered examples: 51
  Class Relative frequency
    1  0.922
    0  0.078

Rule 26
Predicted class is '+'1'

IF Visibility (mi) > 5.5 AND Broken Ceiling > 27.5 AND Wind Speed_0 ≤ 9.5 AND Scattered Ceiling_0 ≤ 37.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = '-'1'

Results on training data

Covered examples: 28
  Class Estimated probability
    1  0.917
    0  0.083

Results on test data

Covered examples: 4
  Class Relative frequency
    1  1.000
    0  0.000

Rule 27
Predicted class is '+'1'

IF Visibility (mi) > 5.5 AND Broken Ceiling ≤ 110 AND Wind Speed_0 > 9.5 AND Scattered Ceiling_0 ≤ 37.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = '-'1'

Results on training data

Covered examples: 11
  Class Estimated probability
    1  0.625
    0  0.375

Results on test data

Covered examples: 8
  Class Relative frequency
    1  0.750
    0  0.250

Rule 28
Predicted class is '-'1'

IF Visibility (mi) > 5.5 AND Broken Ceiling > 110 AND Wind Speed_0 > 9.5 AND Scattered Ceiling_0 ≤ 37.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = '-'1'

Results on training data

Covered examples: 12
  Class Estimated probability
    1  0.111
    0  0.889

Results on test data

Covered examples: 0
  Class Relative frequency
    1  1.000
    0  0.000

Rule 29
Predicted class is '+'1'

IF Visibility (mi) > 5.5 AND Broken Ceiling > 27.5 AND Scattered Ceiling_0 > 37.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 ≤ 12.5 THEN predicted = '-'1'

Results on training data

Covered examples: 38
  Class Estimated probability
    1  0.095
    0  0.905

Results on test data

Covered examples: 18
  Class Relative frequency
    1  0.944
    0  0.056

Rule 30
Predicted class is '-'1'

IF Visibility (mi) > 5.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 > 12.5 AND Overcast Ceiling_2 ≤ 32.5 THEN predicted = '-'1'

Results on training data

Covered examples: 41
  Class Estimated probability
    1  0.143
    0  0.857

Results on test data

Covered examples: 25
  Class Relative frequency
    1  0.000
    0  1.000
### Rule 21
Predicted class is '-1'

**IF** Wind Speed ≤ 6.5 AND Visibility (mi) > 5.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 > 12.5 AND Overcast Ceiling_2 > 32.5 **THEN** predicted = '-1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covered examples:</strong></td>
<td><strong>Covered examples:</strong></td>
</tr>
<tr>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>0.867</td>
<td>0.909</td>
</tr>
<tr>
<td>0.133</td>
<td>0.091</td>
</tr>
</tbody>
</table>

### Rule 32
Predicted class is '-1'

**IF** Wind Speed > 6.5 AND Visibility (mi) > 5.5 AND Wind Speed_1 ≤ 34.5 AND Overcast Ceiling_1 = 12.5 AND Overcast Ceiling_2 > 32.5 **THEN** predicted = '-1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covered examples:</strong></td>
<td><strong>Covered examples:</strong></td>
</tr>
<tr>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>0.706</td>
<td>0.909</td>
</tr>
<tr>
<td>0.294</td>
<td>0.091</td>
</tr>
</tbody>
</table>

### Rule 33
Predicted class is '-1'

**IF** Visibility (mi) > 5.5 AND Wind Speed_1 > 34.5 **THEN** predicted = '-1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covered examples:</strong></td>
<td><strong>Covered examples:</strong></td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>0.800</td>
<td>0.667</td>
</tr>
<tr>
<td>0.200</td>
<td>0.333</td>
</tr>
</tbody>
</table>
I.2 AAR 52 at 0800

Rule 1
Predicted class is '1'
IF Visibility (mi)_0 ≤ 5.5 AND Overcast Ceiling_1 ≤ 2.5 AND Year ≤ 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 93
- Class
  - -1: 0.231
  - 1: 0.769

Results on test data
- Covered examples: 28
- Class
  - -1: 0.250
  - 1: 0.750

Rule 2
Predicted class is '1'
IF Visibility (mi)_0 ≤ 5.5 AND Overcast Ceiling_1 ≤ 2.5 AND Year > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 15
- Class
  - -1: 0.167
  - 1: 0.833

Results on test data
- Covered examples: 14
- Class
  - -1: 0.214
  - 1: 0.786

Rule 3
Predicted class is '1'
IF Visibility (mi)_0 ≤ 5.5 AND Overcast Ceiling_1 > 2.5 AND Overcast Ceiling_1 ≤ 90 THEN predicted = '1'

Results on training data
- Covered examples: 118
- Class
  - -1: 0.100
  - 1: 0.900

Results on test data
- Covered examples: 54
- Class
  - -1: 0.130
  - 1: 0.870

Rule 4
Predicted class is '1'
IF Visibility (mi)_0 ≤ 5.5 AND Overcast Ceiling_1 > 90 THEN predicted = '1'

Results on training data
- Covered examples: 17
- Class
  - -1: 0.222
  - 1: 0.778

Results on test data
- Covered examples: 3
- Class
  - -1: 0.333
  - 1: 0.667

Rule 5
Predicted class is '-1'
IF Visibility (mi)_0 ≤ 4.5 AND Scattered Ceiling ≤ 42.5 AND Visibility (mi)_0 > 5.5 AND Overcast Ceiling_0 ≥ 6 AND Broken Ceiling_0 ≤ 20 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 91
- Class
  - -1: 0.647
  - 1: 0.353

Results on test data
- Covered examples: 31
- Class
  - -1: 0.452
  - 1: 0.548
Rule 6
Predicted class is '1'
IF Visibility (m) ≤ 4.5 AND Scattered Ceiling ≤ 42.5 AND Cross Winds ≤ 0.5 AND Wind Speed_0 ≤ 11.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 6 AND Broken Ceiling_0 = 20 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data

Covered examples: 10
Class Estimated probability
1 0.806
-1 0.200

Results on test data

Covered examples: 7
Class Relative frequency
1 0.714
-1 0.286

Rule 7
Predicted class is '-1'
IF Visibility (m) ≤ 4.5 AND Scattered Ceiling ≤ 42.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 11.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 6 AND Broken Ceiling_0 = 20 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data

Covered examples: 35
Class Estimated probability
1 0.294
-1 0.706

Results on test data

Covered examples: 22
Class Relative frequency
1 0.682
-1 0.318

Rule 8
Predicted class is '-1'
IF Visibility (m) ≤ 4.5 AND Scattered Ceiling ≤ 42.5 AND Wind Speed_0 > 11.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 6 AND Broken Ceiling_0 = 20 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data

Covered examples: 10
Class Estimated probability
1 0.333
-1 0.667

Results on test data

Covered examples: 8
Class Relative frequency
1 0.625
-1 0.375

Rule 9
Predicted class is '-1'
IF Visibility (m) ≤ 4.5 AND Scattered Ceiling > 42.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 6 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data

Covered examples: 9
Class Estimated probability
1 0.500
-1 0.500

Results on test data

Covered examples: 0
Class Relative frequency
1 0.000
-1 0.000

Rule 10
Predicted class is '1'
IF Visibility (m) ≤ 4.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 > 6 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data

Covered examples: 27
Class Estimated probability
1 0.933
-1 0.067

Results on test data

Covered examples: 10
Class Relative frequency
1 0.900
-1 0.100

Rule 11
Predicted class is '-1'
IF Wind Speed ≤ 5.5 AND Visibility (m) AND Cross Winds ≤ 0.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_0 ≤ 55 THEN predicted = '-1'

Results on training data

Covered examples: 46
Class Estimated probability
1 0.174
-1 0.826

Results on test data

Covered examples: 15
Class Relative frequency
1 0.933
-1 0.067
Rule 12
Predicted class is '1'.

IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Cross Winds ≤ 0.5 AND Visibility (mi) ≤ 5.5 AND Overcast Ceiling ≥ 7.5 AND Rain (Y/N) ≤ 0.5 AND Scattered Ceiling ≤ 55 AND Broken Ceiling ≤ 7.5 AND Year ≤ 5.5 THEN predicted = '1'.

Results on training data
Covered examples: 82
-1 0.974
1 0.026

Results on test data
Covered examples: 42
-1 0.857
1 0.143

Rule 13
Predicted class is '1'.

IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Cross Winds ≤ 0.5 AND Wind Speed ≤ 24.5 AND Visibility (mi) ≤ 5.5 AND Overcast Ceiling ≤ 7.5 AND Rain (Y/N) ≤ 0.5 AND Scattered Ceiling ≤ 55 AND Broken Ceiling ≤ 7.5 AND Year > 5.5 THEN predicted = '1'.

Results on training data
Covered examples: 15
-1 0.800
1 0.200

Results on test data
Covered examples: 8
-1 0.875
1 0.125

Rule 14
Predicted class is '1'.

IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Cross Winds ≤ 0.5 AND Wind Speed ≤ 24.5 AND Visibility (mi) ≤ 5.5 AND Overcast Ceiling ≤ 7.5 AND Rain (Y/N) ≤ 0.5 AND Scattered Ceiling ≤ 55 AND Broken Ceiling ≤ 7.5 AND Year > 5.5 THEN predicted = '1'.

Results on training data
Covered examples: 8
-1 0.200
1 0.800

Results on test data
Covered examples: 3
-1 1.000
1 0.000

Rule 15
Predicted class is '1'.

IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Broken Ceiling ≤ 25 AND Cross Winds ≤ 0.5 AND Visibility (mi) ≤ 5.5 AND Overcast Ceiling ≤ 7.5 AND Rain (Y/N) ≤ 0.5 AND Scattered Ceiling ≤ 55 AND Broken Ceiling ≤ 7.5 AND Broken Ceiling > 105 THEN predicted = '1'.

Results on training data
Covered examples: 18
-1 0.846
1 0.154

Results on test data
Covered examples: 15
-1 0.733
1 0.267

Rule 16
Predicted class is '1'.

IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Broken Ceiling > 25 AND Cross Winds ≤ 0.5 AND Visibility (mi) ≤ 5.5 AND Overcast Ceiling ≤ 7.5 AND Rain (Y/N) ≤ 0.5 AND Scattered Ceiling ≤ 55 AND Broken Ceiling > 7.5 AND Broken Ceiling ≥ 105 THEN predicted = '1'.

Results on training data
Covered examples: 25
-1 0.667
1 0.333

Results on test data
Covered examples: 8
-1 0.625
1 0.375
Rule 17
Predicted class is ‘-1’
IF Wind Speed > 5.5 AND Visibility (mi) > 4.5 AND Cross Winds ≤ 0.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N) _2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 105 THEN predicted = ‘-1’

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.854</td>
</tr>
<tr>
<td>1</td>
<td>0.154</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Rule 18
Predicted class is ‘-1’
IF Visibility (mi) > 4.5 AND Cross Winds ≤ 0.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N) _2 ≤ 0.5 AND Scattered Ceiling_2 > 55 THEN predicted = ‘-1’

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.992</td>
</tr>
<tr>
<td>1</td>
<td>0.108</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.969</td>
</tr>
<tr>
<td>1</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Rule 19
Predicted class is ‘-1’
IF Visibility (mi) > 4.5 AND Cross Winds > 0.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N) _2 ≤ 0.5 AND Year ≤ 3.5 THEN predicted = ‘-1’

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.938</td>
</tr>
<tr>
<td>1</td>
<td>0.162</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.891</td>
</tr>
<tr>
<td>1</td>
<td>0.109</td>
</tr>
</tbody>
</table>

Rule 20
Predicted class is ‘-1’
IF Visibility (mi) > 4.5 AND Cross Winds > 0.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 ≤ 4.5 AND Rain (Y/N) _2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Year > 3.5 THEN predicted = ‘-1’

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.667</td>
</tr>
<tr>
<td>1</td>
<td>0.333</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.900</td>
</tr>
<tr>
<td>1</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Rule 21
Predicted class is ‘-1’
IF Visibility (mi) > 4.5 AND Visibility (mi) ≤ 5.5 AND Cross Winds > 0.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Rain (Y/N) _2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = ‘-1’

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.750</td>
</tr>
<tr>
<td>1</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.856</td>
</tr>
<tr>
<td>1</td>
<td>0.444</td>
</tr>
</tbody>
</table>
Rule 22
Predicted class is '-1'.

IF Visibility (m) > 5.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 9 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Scattered Ceiling_0 ≤ 65 AND Wind Speed_1 ≤ 21 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Wind Speed_2 ≤ 6.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 13
  - Class Estimated probability
    - -1: 0.500
    - 1: 0.500

Results on test data
- Covered examples: 6
  - Class Relative frequency
    - -1: 0.833
    - 1: 0.167

Rule 23
Predicted class is '1'.

IF Visibility (m) > 5.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 9 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Scattered Ceiling_0 ≤ 65 AND Wind Speed_1 ≤ 21 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Wind Speed_2 ≤ 6.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = '1'.

Results on training data
- Covered examples: 7
  - Class Estimated probability
    - -1: 0.750
    - 1: 0.250

Results on test data
- Covered examples: 2
  - Class Relative frequency
    - -1: 0.000
    - 1: 1.000

Rule 24
Predicted class is '-1'.

IF Visibility (m) > 5.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 9 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Wind Speed_1 ≤ 21 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 37
  - Class Estimated probability
    - -1: 0.100
    - 1: 0.900

Results on test data
- Covered examples: 9
  - Class Relative frequency
    - -1: 0.091
    - 1: 0.909

Rule 25
Predicted class is '-1'.

IF Visibility (m) > 5.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 9 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Wind Speed_1 ≤ 21 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 30
  - Class Estimated probability
    - -1: 0.333
    - 1: 0.667

Results on test data
- Covered examples: 17
  - Class Relative frequency
    - -1: 0.529
    - 1: 0.471

Rule 26
Predicted class is '-1'.

IF Visibility (m) > 5.5 AND Cross Winds > 0.5 AND Visibility (m)_0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Wind Speed_1 ≤ 21 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 ≤ 175 AND Year > 3.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 29
  - Class Estimated probability
    - -1: 0.238
    - 1: 0.762

Results on test data
- Covered examples: 7
  - Class Relative frequency
    - -1: 0.286
    - 1: 0.714
Rule 27
Predicted class is -1
IF Visibility (m) > 4.5 AND Cross Winds > 0.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Few Ceiling_1 ≤ 20 AND Wind Speed_2 > 4.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 17 AND Year > 3.5 THEN predicted = -1

Results on training data
- Covered examples: 20
  - Class: Estimated probability
    - 1: 0.750
    - -1: 0.250

Results on test data
- Covered examples: 4
  - Class: Relative frequency
    - 1: 0.500
    - -1: 0.500

Rule 28
Predicted class is -1
IF Visibility (m) > 4.5 AND Cross Winds > 0.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Few Ceiling_1 > 20 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Year > 3.5 THEN predicted = -1

Results on training data
- Covered examples: 26
  - Class: Estimated probability
    - -1: 0.807
    - 1: 0.193

Results on test data
- Covered examples: 17
  - Class: Relative frequency
    - -1: 0.824
    - 1: 0.176

Rule 29
Predicted class is -1
IF Visibility (m) > 4.5 AND Cross Winds > 0.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 225 AND Year > 3.5 THEN predicted = -1

Results on training data
- Covered examples: 34
  - Class: Estimated probability
    - -1: 0.714
    - 1: 0.286

Results on test data
- Covered examples: 11
  - Class: Relative frequency
    - -1: 0.818
    - 1: 0.182

Rule 30
Predicted class is -1
IF Visibility (m) > 4.5 AND Cross Winds > 0.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 ≤ 7.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 > 225 AND Year > 3.5 THEN predicted = -1

Results on training data
- Covered examples: 19
  - Class: Estimated probability
    - -1: 0.833
    - 1: 0.167

Results on test data
- Covered examples: 11
  - Class: Relative frequency
    - -1: 0.777
    - 1: 0.223

Rule 31
Predicted class is -1
IF Visibility (m) > 4.5 AND Wind Speed_0 ≤ 6.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 > 7.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = -1

Results on training data
- Covered examples: 15
  - Class: Estimated probability
    - -1: 0.800
    - 1: 0.200

Results on test data
- Covered examples: 11
  - Class: Relative frequency
    - -1: 0.545
    - 1: 0.455

Rule 32
Predicted class is -1
IF Visibility (m) > 4.5 AND Wind Speed_0 ≤ 6.5 AND Visibility (m)0 > 5.5 AND Overcast Ceiling_0 > 7.5 AND Cross Winds_0 ≤ 0.5 AND Wind Speed_1 ≤ 11.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = -1

Results on training data
- Covered examples: 13
  - Class: Estimated probability
    - -1: 0.889
    - 1: 0.111

Results on test data
- Covered examples: 6
  - Class: Relative frequency
    - -1: 0.167
    - 1: 0.833
Rule 33
Predicted class is '1'
IF Visibility (mi) > 4.5 AND Wind Speed_0 > 6.5 AND Visibility (mi)_0 > 5.5 AND Overcast Ceiling_0 > 7.5 AND Cross Winds_0 > 0.5 AND Wind Speed_1 ≤ 11.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data
Covered examples: 14
Class Estimated probability
-1 0.455
1 0.545

Results on test data
Covered examples: 7
Class Relative frequency
-1 0.286
1 0.714

Rule 34
Predicted class is '1'
IF Visibility (mi) > 4.5 AND Wind Speed_0 > 6.5 AND Visibility (mi)_0 > 5.5 AND Overcast Ceiling_0 > 7.5 AND Wind Speed_1 > 11.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data
Covered examples: 34
Class Estimated probability
-1 0.476
1 0.524

Results on test data
Covered examples: 7
Class Relative frequency
-1 0.571
1 0.429

Rule 35
Predicted class is '1'
IF Scattered Ceiling ≤ 30 AND Broken Ceiling ≤ 25 AND Visibility (mi)_0 > 5.5 AND Scattered Ceiling_0 ≤ 15 AND Rain (Y/N)_2 > 0.5 AND Scattered Ceiling_2 ≤ 27.5 THEN predicted = '1'

Results on training data
Covered examples: 36
Class Estimated probability
-1 0.176
1 0.824

Results on test data
Covered examples: 21
Class Relative frequency
-1 0.286
1 0.714

Rule 36
Predicted class is '1'
IF Visibility (mi) ≤ 4.5 AND Scattered Ceiling ≤ 30 AND Broken Ceiling ≤ 25 AND Visibility (mi)_0 > 5.5 AND Scattered Ceiling_0 > 15 AND Rain (Y/N)_2 > 0.5 AND Scattered Ceiling_2 ≤ 27.5 THEN predicted = '1'

Results on training data
Covered examples: 11
Class Estimated probability
-1 0.286
1 0.714

Results on test data
Covered examples: 8
Class Relative frequency
-1 0.625
1 0.375

Rule 37
Predicted class is '1'
IF Visibility (mi) > 4.5 AND Scattered Ceiling ≤ 30 AND Broken Ceiling ≤ 25 AND Visibility (mi)_0 > 5.5 AND Scattered Ceiling_0 > 15 AND Rain (Y/N)_2 > 0.5 AND Scattered Ceiling_2 ≤ 27.5 THEN predicted = '1'

Results on training data
Covered examples: 8
Class Estimated probability
-1 0.800
1 0.200

Results on test data
Covered examples: 5
Class Relative frequency
-1 0.800
1 0.200

Rule 38
Predicted class is '1'
IF Scattered Ceiling > 30 AND Broken Ceiling ≤ 25 AND Visibility (mi)_0 > 5.5 AND Rain (Y/N)_2 > 0.5 AND Scattered Ceiling_2 > 27.5 THEN predicted = '1'

Results on training data
Covered examples: 15
Class Estimated probability
-1 0.333
1 0.667

Results on test data
Covered examples: 5
Class Relative frequency
-1 0.800
1 0.200

168
Rule 39
Predicted class is ‘1’
IF Broken Ceiling > 25 AND Visibility (mi) 0 > 5.5 AND Rain (Y/N) > 0.5 AND Overcast Ceiling 0 2 ≤ 8.5 AND Scattered Ceiling 0 2 ≤ 27.5 THEN predicted = ‘1’

Results on training data
Covered examples: 20
Class Estimated probability
-1 0.333
1 0.667

Results on test data
Covered examples: 7
Class Relative frequency
-1 0.714
1 0.286

Rule 40
Predicted class is ‘-1’
IF Broken Ceiling > 25 AND Visibility (mi) 0 > 5.5 AND Rain (Y/N) > 0.5 AND Overcast Ceiling 0 2 > 8.5 AND Scattered Ceiling 0 2 ≤ 27.5 THEN predicted = ‘-1’

Results on training data
Covered examples: 18
Class Estimated probability
-1 0.615
1 0.385

Results on test data
Covered examples: 5
Class Relative frequency
-1 1.000
1 0.000

Rule 41
Predicted class is ‘1’
IF Visibility (mi) ≤ 5.5 AND Visibility (mi) 0 > 5.5 AND Rain (Y/N) > 0.5 AND Scattered Ceiling 0 2 > 27.5 THEN predicted = ‘1’

Results on training data
Covered examples: 18
Class Estimated probability
-1 0.467
1 0.533

Results on test data
Covered examples: 4
Class Relative frequency
-1 0.750
1 0.250

Rule 42
Predicted class is ‘1’
IF Visibility (mi) > 5.5 AND Visibility (mi) 0 > 5.5 AND Rain (Y/N) > 0.5 AND Scattered Ceiling 0 2 > 27.5 THEN predicted = ‘1’

Results on training data
Covered examples: 7
Class Estimated probability
-1 0.250
1 0.750

Results on test data
Covered examples: 5
Class Relative frequency
-1 0.600
1 0.400
1.3 AAR 48 at 1200

Rule 1
Predicted class is ‘-1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 11 AND Scattered Ceiling_1 ≤ 27.5 AND Broken Ceiling_1 ≤ 37.5 AND Wind Speed_2 ≤ 12.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 55
  - Class Estimated probability
    -1 0.571
    1 0.429

Results on test data
- Covered examples: 18
  - Class Relative frequency
    -1 0.722
    1 0.278

Rule 2
Predicted class is ‘-1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 11 AND Scattered Ceiling_1 ≤ 27.5 AND Broken Ceiling_1 > 37.5 AND Wind Speed_2 ≤ 12.5 AND Broken Ceiling_2 ≤ 85 THEN predicted = ‘-1’

Results on training data
- Covered examples: 17
  - Class Estimated probability
    -1 0.899
    1 0.111

Results on test data
- Covered examples: 3
  - Class Relative frequency
    -1 1.000
    1 0.000

Rule 3
Predicted class is ‘-1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 11 AND Scattered Ceiling_1 ≤ 27.5 AND Broken Ceiling_1 > 37.5 AND Wind Speed_2 ≤ 12.5 AND Broken Ceiling_2 > 85 THEN predicted = ‘-1’

Results on training data
- Covered examples: 17
  - Class Estimated probability
    -1 0.615
    1 0.385

Results on test data
- Covered examples: 4
  - Class Relative frequency
    -1 1.000
    1 0.000

Rule 4
Predicted class is ‘-1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Overcast Ceiling > 11 AND Scattered Ceiling_1 ≤ 27.5 AND Wind Speed_2 ≤ 12.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 9
  - Class Estimated probability
    -1 0.750
    1 0.250

Results on test data
- Covered examples: 4
  - Class Relative frequency
    -1 0.500
    1 0.500

Rule 5
Predicted class is ‘-1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Wind Speed_0 ≤ 16 AND Scattered Ceiling_1 ≤ 27.5 AND Wind Speed_2 > 12.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 14
  - Class Estimated probability
    -1 0.500
    1 0.500

Results on test data
- Covered examples: 5
  - Class Relative frequency
    -1 0.200
    1 0.800

Rule 6
Predicted class is ‘1’
IF Visibility (mi) ≤ 3.5 AND Rain (Y/N) ≤ 0.5 AND Wind Speed_0 > 16 AND Scattered Ceiling_1 ≤ 27.5 AND Wind Speed_2 > 12.5 THEN predicted = ‘1’

Results on training data
- Covered examples: 8
  - Class Estimated probability
    -1 0.250
    1 0.750

Results on test data
- Covered examples: 3
  - Class Relative frequency
    -1 0.000
    1 1.000
Rule 7
Predicted class is '-1'
IF Wind Speed $\leq$ 3.5 AND Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $\leq$ 0.5 AND Mist (Y/N)_0 $\leq$ 0.5 AND Scattered Ceiling_1 $>$ 27.5 THEN predicted = '-1'

Results on training data
- Covered examples: 27
  - Class: Estimated probability
    - 1: 0.632
    - 1: 0.368

Results on test data
- Covered examples: 16
  - Class: Relative frequency
    - 1: 0.613
    - 1: 0.387

Rule 8
Predicted class is '-1'
IF Wind Speed $>$ 3.5 AND Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $\leq$ 0.5 AND Mist (Y/N)_0 $\leq$ 0.5 AND Scattered Ceiling_1 $>$ 27.5 THEN predicted = '-1'

Results on training data
- Covered examples: 45
  - Class: Estimated probability
    - 1: 0.852
    - 1: 0.148

Results on test data
- Covered examples: 17
  - Class: Relative frequency
    - 1: 0.765
    - 1: 0.235

Rule 9
Predicted class is '-1'
IF Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $\leq$ 0.5 AND Mist (Y/N)_0 $>$ 0.5 AND Scattered Ceiling_1 $>$ 27.5 THEN predicted = '-1'

Results on training data
- Covered examples: 11
  - Class: Estimated probability
    - 1: 0.500
    - 1: 0.500

Results on test data
- Covered examples: 3
  - Class: Relative frequency
    - 1: 0.000
    - 1: 1.000

Rule 10
Predicted class is '1'
IF Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $>$ 0.5 AND Overcast Ceiling_0 $\leq$ 75 AND Scattered Ceiling_2 $\leq$ 12.5 AND Broken Ceiling_2 $\leq$ 95 THEN predicted = '1'

Results on training data
- Covered examples: 74
  - Class: Estimated probability
    - 1: 0.342
    - 1: 0.658

Results on test data
- Covered examples: 32
  - Class: Relative frequency
    - 1: 0.375
    - 1: 0.625

Rule 11
Predicted class is '-1'
IF Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $>$ 0.5 AND Overcast Ceiling_0 $>$ 75 AND Scattered Ceiling_2 $\leq$ 12.5 AND Broken Ceiling_2 $\geq$ 95 THEN predicted = '-1'

Results on training data
- Covered examples: 8
  - Class: Estimated probability
    - 1: 0.800
    - 1: 0.200

Results on test data
- Covered examples: 3
  - Class: Relative frequency
    - 1: 0.667
    - 1: 0.333

Rule 12
Predicted class is '1'
IF Visibility (m) $\leq$ 3.5 AND Rain (Y/N) $>$ 0.5 AND Scattered Ceiling_2 $\geq$ 12.5 AND Broken Ceiling_2 $\leq$ 85 THEN predicted = '1'

Results on training data
- Covered examples: 19
  - Class: Estimated probability
    - 1: 0.375
    - 1: 0.625

Results on test data
- Covered examples: 3
  - Class: Relative frequency
    - 1: 0.667
    - 1: 0.333
Rule 13
Predicted class is '1'
IF Visibility (m) ≤ 3.5 AND Rain (Y/N) > 0.5 AND Broken Ceiling_2 > 85 THEN predicted = '1'

Results on training data
Covered examples: 11

Class       Estimated probability
-1          0.429
1           0.571

Results on test data
Covered examples: 1
Class       Relative frequency
-1          1.000
1           0.000

Rule 14
Predicted class is '-1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_0 ≤ 6 AND Wind Speed_1 ≤ 8.5 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 ≤ 5.5 AND Showers (Y/N)_2 ≤ 6.5 THEN predicted = '-1'

Results on training data
Covered examples: 46

Class       Estimated probability
-1          0.085
1           0.115

Results on test data
Covered examples: 24

Class       Relative frequency
-1          1.000
1           0.000

Rule 15
Predicted class is '-1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_0 ≤ 6 AND Wind Speed_1 > 8.5 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 ≤ 5.5 AND Showers (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
Covered examples: 20

Class       Estimated probability
-1          0.917
1           0.083

Results on test data
Covered examples: 4

Class       Relative frequency
-1          0.500
1           0.500

Rule 16
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_0 > 6 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 ≤ 5.5 AND Showers (Y/N)_2 ≤ 0.5 THEN predicted = '1'

Results on training data
Covered examples: 23

Class       Estimated probability
-1          0.917
1           0.083

Results on test data
Covered examples: 5

Class       Relative frequency
-1          1.000
1           0.000

Rule 17
Predicted class is '-1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_1 ≤ 41.5 AND Wind Speed_2 > 9.5 AND Wind Speed_1 ≤ 17.5 AND Showers (Y/N)_1 ≤ 5.5 AND Broken Ceiling_2 ≤ 12.5 THEN predicted = '-1'

Results on training data
Covered examples: 179

Class       Estimated probability
-1          0.946
1           0.054

Results on test data
Covered examples: 63

Class       Relative frequency
-1          0.873
1           0.127

Rule 18
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Wind Speed_0 ≤ 6.5 AND Wind Speed_1 ≤ 8.5 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 > 5.5 AND Wind Speed_2 ≤ 17.5 AND Showers (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 12.5 THEN predicted = '1'

Results on training data
Covered examples: 25

Class       Estimated probability
-1          0.909
1           0.091

Results on test data
Covered examples: 12

Class       Relative frequency
-1          0.917
1           0.083
Rule 19
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Wind Speed_0 > 6.5 AND Wind Speed_1 ≤ 8.5 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 > 5.5 AND Wind Speed_2 ≤ 17.5 AND Showers (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 12.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 15
  - Class Estimated probability:
    - -1 0.600
    - 1 0.400

Results on test data
- Covered examples: 13
  - Class Relative frequency:
    - -1 1.000
    - 1 0.000

Rule 20
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Wind Speed_1 > 8.5 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 > 5.5 AND Wind Speed_2 ≤ 17.5 AND Showers (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 AND Broken Ceiling_2 > 12.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 58
  - Class Estimated probability:
    - -1 0.879
    - 1 0.121

Results on test data
- Covered examples: 21
  - Class Relative frequency:
    - -1 0.905
    - 1 0.095

Rule 21
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 > 5.5 AND Wind Speed_2 ≤ 17.5 AND Showers (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 THEN predicted = '-1'.

Results on training data
- Covered examples: 30
  - Class Estimated probability:
    - -1 0.938
    - 1 0.063

Results on test data
- Covered examples: 17
  - Class Relative frequency:
    - -1 1.000
    - 1 0.000

Rule 22
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_1 ≤ 42.5 AND Wind Speed_2 > 17.5 AND Showers (Y/N)_2 ≤ 0.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 58
  - Class Estimated probability:
    - -1 0.780
    - 1 0.220

Results on test data
- Covered examples: 32
  - Class Relative frequency:
    - -1 0.781
    - 1 0.219

Rule 23
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Scattered Ceiling_1 > 42.5 AND Showers (Y/N)_2 ≤ 0.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 287
  - Class Estimated probability:
    - -1 0.972
    - 1 0.028

Results on test data
- Covered examples: 144
  - Class Relative frequency:
    - -1 0.986
    - 1 0.014

Rule 24
Predicted class is '-1'.
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Cross Winds_0 ≤ 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = '-1'.

Results on training data
- Covered examples: 26
  - Class Estimated probability:
    - -1 0.875
    - 1 0.125

Results on test data
- Covered examples: 10
  - Class Relative frequency:
    - -1 1.000
    - 1 0.000
Rule 25
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling_0 ≤ 30 AND Cross Winds_0 > 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = '1'

Results on training data
Covered examples: 14
-1 0.700
1 0.300

Results on test data
Covered examples: 4
-1 0.500
1 0.500

Rule 26
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling_0 > 30 AND Cross Winds_0 > 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = '1'

Results on training data
Covered examples: 7
-1 0.750
1 0.250

Results on test data
Covered examples: 9
-1 0.778
1 0.222

Rule 27
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling > 2 AND Overcast Ceiling ≤ 13.5 THEN predicted = '1'

Results on training data
Covered examples: 27
-1 0.700
1 0.300

Results on test data
Covered examples: 13
-1 0.615
1 0.385

Rule 28
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling > 13.5 AND Visibility (m)_2 ≤ 4.5 THEN predicted = '1'

Results on training data
Covered examples: 27
-1 0.615
1 0.385

Results on test data
Covered examples: 11
-1 0.545
1 0.455

Rule 29
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling > 13.5 AND Overcast Ceiling_1 ≤ 22.5 AND Visibility (m)_2 > 4.5 THEN predicted = '1'

Results on training data
Covered examples: 66
-1 0.842
1 0.158

Results on test data
Covered examples: 35
-1 0.857
1 0.143

Rule 30
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling > 13.5 AND Wind Speed_1 ≤ 12.5 AND Overcast Ceiling_1 = 22.5 AND Visibility (m)_2 > 4.5 THEN predicted = '1'

Results on training data
Covered examples: 34
-1 0.857
1 0.143

Results on test data
Covered examples: 12
-1 0.833
1 0.167
Rule 31

Predicted class is '1'

IF Visibility (m) > 3.5 AND Overcast Ceiling > 13.5 AND Wind Speed_1 > 12.5 AND Overcast Ceiling_1 > 22.5 AND Visibility (m)_2 > 4.5 THEN predicted = '1'

Results on training data

<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.444</td>
</tr>
<tr>
<td>1</td>
<td>0.556</td>
</tr>
</tbody>
</table>

Results on test data

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.833</td>
</tr>
<tr>
<td>1</td>
<td>0.167</td>
</tr>
</tbody>
</table>
I.4  AAR 52 at 1200

Rule 1
Predicted class is '1'.
IF Visibility (mi) ≤ 3.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year ≤ 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 15
  - Class Estimated probability:
    - 1: 0.333
    - -1: 0.667

Results on test data
- Covered examples: 5
  - Class Estimated probability:
    - 1: 0.200
    - -1: 0.800

Rule 2
Predicted class is '1'.
IF Visibility (mi) > 3.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year ≤ 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 15
  - Class Estimated probability:
    - 1: 0.182
    - -1: 0.818

Results on test data
- Covered examples: 11
  - Class Estimated probability:
    - 1: 0.273
    - -1: 0.727

Rule 3
Predicted class is '1'.
IF Wind Speed ≤ 4.5 AND Snow (Y/N) ≤ 0.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year > 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 26
  - Class Estimated probability:
    - 1: 0.200
    - -1: 0.800

Results on test data
- Covered examples: 10
  - Class Estimated probability:
    - 1: 0.200
    - -1: 0.800

Rule 4
Predicted class is '1'.
IF Wind Speed > 4.5 AND Wind Speed ≤ 5.5 AND Snow (Y/N) ≤ 0.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year > 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 21
  - Class Estimated probability:
    - 1: 0.333
    - -1: 0.667

Results on test data
- Covered examples: 8
  - Class Estimated probability:
    - 1: 0.250
    - -1: 0.750

Rule 5
Predicted class is '1'.
IF Wind Speed > 5.5 AND Snow (Y/N) ≤ 0.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year > 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 63
  - Class Estimated probability:
    - 1: 0.100
    - -1: 0.900

Results on test data
- Covered examples: 25
  - Class Estimated probability:
    - 1: 0.080
    - -1: 0.920

Rule 6
Predicted class is '1'.
IF Snow (Y/N) > 0.5 AND Broken Ceiling ≤ 16.5 AND Visibility (mi)₂ ≤ 5.5 AND Year > 2.5 THEN predicted = '1'

Results on training data
- Covered examples: 7
  - Class Estimated probability:
    - 1: 0.250
    - -1: 0.750

Results on test data
- Covered examples: 3
  - Class Estimated probability:
    - 1: 0.333
    - -1: 0.667
Rule 7
Predicted class is '1'
IF Broken Ceiling > 16.5 AND Mist (Y/N)_1 ≤ 0.5 AND Visibility (mi)_2 ≤ 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 24
  - Class 1: Estimated probability = 0.400
  - Class 2: Estimated probability = 0.600

Results on test data
- Covered examples: 20
  - Class 1: Estimated relative frequency = 0.450
  - Class 2: Estimated relative frequency = 0.550

Rule 8
Predicted class is '1'
IF Broken Ceiling > 16.5 AND Mist (Y/N)_1 > 0.5 AND Visibility (mi)_2 ≤ 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 18
  - Class 1: Estimated probability = 0.756

Results on test data
- Covered examples: 7
  - Class 1: Estimated relative frequency = 0.714

Rule 9
Predicted class is '1'
IF Wind Speed ≤ 3.5 AND Visibility (mi) ≤ 3.5 AND Fog (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 1.5 AND Wind Speed_1 ≤ 6.5 AND Broken Ceiling_1 ≤ 125 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 21
  - Class 1: Estimated probability = 0.500

Results on test data
- Covered examples: 3
  - Class 1: Estimated relative frequency = 0.333

Rule 10
Predicted class is '1'
IF Wind Speed > 3.5 AND Visibility (mi) ≤ 3.5 AND Fog (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 1.5 AND Wind Speed_1 ≤ 6.5 AND Broken Ceiling_1 ≤ 125 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 8
  - Class 1: Estimated probability = 0.800

Results on test data
- Covered examples: 5
  - Class 1: Estimated relative frequency = 0.200

Rule 11
Predicted class is '1'
IF Visibility (mi) ≤ 3.5 AND Fog (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 1.5 AND Wind Speed_1 > 6.5 AND Broken Ceiling_1 ≤ 125 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 78
  - Class 1: Estimated probability = 0.511

Results on test data
- Covered examples: 31
  - Class 1: Estimated relative frequency = 0.516

Rule 12
Predicted class is '1'
IF Visibility (mi) ≤ 3.5 AND Fog (Y/N) ≤ 0.5 AND Overcast Ceiling ≤ 1.5 AND Broken Ceiling_1 > 125 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 17
  - Class 1: Estimated probability = 0.286

Results on test data
- Covered examples: 5
  - Class 1: Estimated relative frequency = 0.600
Rule 13
Predicted class is '1'
IF Visibility (m) ≤ 3.5 AND Fog (Y/N) > 0.5 AND Overcast Ceiling ≤ 1.5 AND Visibility (m)_2 > 5.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.625</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 13
  - Class Estimated probability: -1 0.375, 1 0.625

Results on test data:
- Covered examples: 5
  - Class Estimated probability: -1 0.400, 1 0.600

Rule 14
Predicted class is '1'
IF Visibility (m) ≤ 3.5 AND Overcast Ceiling > 1.5 AND Overcast Ceiling_1 ≤ 32.5 AND Visibility (m)_2 > 5.5 AND Scattered Ceiling_2 ≤ 42.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.655</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 57
  - Class Estimated probability: -1 0.345, 1 0.655

Results on test data:
- Covered examples: 28
  - Class Estimated probability: -1 0.179, 1 0.821

Rule 15
Predicted class is '1'
IF Visibility (m) ≤ 3.5 AND Overcast Ceiling > 1.5 AND Overcast Ceiling_1 > 32.5 AND Visibility (m)_2 > 5.5 AND Scattered Ceiling_2 ≤ 42.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.833</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 12
  - Class Estimated probability: -1 0.167, 1 0.833

Results on test data:
- Covered examples: 4
  - Class Estimated probability: -1 0.000, 1 1.000

Rule 16
Predicted class is '1'
IF Visibility (m) ≤ 3.5 AND Overcast Ceiling > 1.5 AND Visibility (m)_2 > 5.5 AND Scattered Ceiling_2 > 42.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.429</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 11
  - Class Estimated probability: -1 0.429, 1 0.571

Results on test data:
- Covered examples: 4
  - Class Estimated probability: -1 0.750, 1 0.250

Rule 17
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 160 AND Cross Winds ≤ 0.5 AND Wind Speed_0 ≤ 6.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 ≤ 8.5 AND Visibility (m)_2 > 5.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 23
  - Class Estimated probability: -1 0.923, 1 0.077

Results on test data:
- Covered examples: 10
  - Class Estimated probability: -1 0.600, 1 0.400

Rule 18
Predicted class is '1'
IF Visibility (m) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 160 AND Cross Winds ≤ 0.5 AND Wind Speed_0 > 6.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 ≤ 8.5 AND Visibility (m)_2 > 5.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Predicted class</th>
<th>Coverage</th>
<th>Class Estimated probability</th>
<th>Class Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.455</td>
</tr>
</tbody>
</table>

Results on training data:
- Covered examples: 16
  - Class Estimated probability: -1 0.545, 1 0.455

Results on test data:
- Covered examples: 7
  - Class Estimated probability: -1 0.571, 1 0.429
Rule 19
Predicted class is '1'.
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 160 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 > 8.5 AND Wind Speed_1 ≤ 23.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'.

Results on training data:
- Covered examples: 65
  - Class Estimated probability:
    - -1 0.735
    - 1 0.265

Results on test data:
- Covered examples: 31
  - Class Relative frequency:
    - -1 0.039
    - 1 0.161

Rule 20
Predicted class is '1'.
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 160 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 > 23.5 AND Visibility (mi)_2 > 5.5 AND Broken Ceiling_2 ≤ 20 THEN predicted = '1'.

Results on training data:
- Covered examples: 21
  - Class Estimated probability:
    - -1 0.941
    - 1 0.059

Results on test data:
- Covered examples: 12
  - Class Relative frequency:
    - -1 0.833
    - 1 0.167

Rule 21
Predicted class is '1'.
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 160 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 > 23.5 AND Visibility (mi)_2 > 5.5 AND Broken Ceiling_2 > 20 THEN predicted = '1'.

Results on training data:
- Covered examples: 14
  - Class Estimated probability:
    - -1 0.727
    - 1 0.273

Results on test data:
- Covered examples: 8
  - Class Relative frequency:
    - -1 0.500
    - 1 0.500

Rule 22
Predicted class is '1'.
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling > 160 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 ≤ 32.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'.

Results on training data:
- Covered examples: 10
  - Class Estimated probability:
    - -1 0.500
    - 1 0.500

Results on test data:
- Covered examples: 1
  - Class Relative frequency:
    - -1 1.000
    - 1 0.000

Rule 23
Predicted class is '1'.
IF Wind Speed ≤ 3.5 AND Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 ≤ 6.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'.

Results on training data:
- Covered examples: 19
  - Class Estimated probability:
    - -1 0.917
    - 1 0.083

Results on test data:
- Covered examples: 7
  - Class Relative frequency:
    - -1 1.000
    - 1 0.000

Rule 24
Predicted class is '1'.
IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 ≤ 6.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'.

Results on training data:
- Covered examples: 20
  - Class Estimated probability:
    - -1 0.727
    - 1 0.273

Results on test data:
- Covered examples: 8
  - Class Relative frequency:
    - -1 0.500
    - 1 0.500
Rule 25
Predicted class is ‘-1’
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 6.5 AND Wind Speed_2 ≤ 7.5 AND Visibility (mi)_2 > 5.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 13
  - Class Estimated probability
    - 1: 0.667
    - 2: 0.333

Results on test data
- Covered examples: 8
  - Class Relative frequency
    - 1: 0.875
    - 2: 0.125

Rule 26
Predicted class is ‘-1’
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 7.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 11
  - Class Estimated probability
    - 1: 0.857
    - 2: 0.143

Results on test data
- Covered examples: 6
  - Class Relative frequency
    - 1: 0.500
    - 2: 0.500

Rule 27
Predicted class is ‘-1’
IF Wind Speed ≤ 7 AND Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 > 6.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 7.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 10
  - Class Estimated probability
    - 1: 0.571
    - 2: 0.429

Results on test data
- Covered examples: 5
  - Class Relative frequency
    - 1: 0.400
    - 2: 0.600

Rule 28
Predicted class is ‘-1’
IF Wind Speed > 7 AND Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling ≤ 2.5 AND Cross Winds > 0.5 AND Wind Speed_0 > 6.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 7.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 7
  - Class Estimated probability
    - 1: 0.750
    - 2: 0.250

Results on test data
- Covered examples: 3
  - Class Relative frequency
    - 1: 0.667
    - 2: 0.333

Rule 29
Predicted class is ‘1’
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling > 2.5 AND Broken Ceiling ≤ 200 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = ‘1’

Results on training data
- Covered examples: 32
  - Class Estimated probability
    - 1: 0.421
    - 2: 0.579

Results on test data
- Covered examples: 4
  - Class Relative frequency
    - 1: 0.750
    - 2: 0.250
Rule 30
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Broken Ceiling > 200 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
- Covered examples: 11
  - Class Estimated probability
    - 1: 0.571
    - 0: 0.429

Results on test data
- Covered examples: 9
  - Class Relative frequency
    - 1: 0.689
    - 0: 0.111

Rule 31
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Cross Winds > 0.5 AND Wind Speed_0 > 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 ≤ 12.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
- Covered examples: 47
  - Class Estimated probability
    - 1: 0.696
    - 0: 0.304

Results on test data
- Covered examples: 16
  - Class Relative frequency
    - 1: 0.688
    - 0: 0.313

Rule 32
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Cross Winds > 0.5 AND Wind Speed_0 > 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 > 12.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 AND Cross Winds_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 17
  - Class Estimated probability
    - 1: 0.500
    - 0: 0.500

Results on test data
- Covered examples: 5
  - Class Relative frequency
    - 1: 0.600
    - 0: 0.400

Rule 33
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Cross Winds > 0.5 AND Wind Speed_0 > 8.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_1 > 12.5 AND Wind Speed_2 ≤ 19 AND Visibility (mi)_2 > 5.5 AND Cross Winds_2 > 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 18
  - Class Estimated probability
    - 1: 0.800
    - 0: 0.200

Results on test data
- Covered examples: 4
  - Class Relative frequency
    - 1: 0.750
    - 0: 0.250

Rule 34
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling ≤ 2 AND Cross Winds > 0.5 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 19 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
- Covered examples: 17
  - Class Estimated probability
    - 1: 0.500
    - 0: 0.500

Results on test data
- Covered examples: 12
  - Class Relative frequency
    - 1: 0.583
    - 0: 0.417

Rule 35
Predicted class is '1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling > 2 AND Overcast Ceiling ≤ 17.5 AND Scattered Ceiling_0 ≤ 32.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 26
  - Class Estimated probability
    - 1: 0.133
    - 0: 0.867

Results on test data
- Covered examples: 7
  - Class Relative frequency
    - 1: 0.714
    - 0: 0.286
Rule 36
Predicted class is '1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling > 17.5 AND Overcast Ceiling ≤ 65 AND Scattered Ceiling_0 ≤ 32.5 AND Wind Speed_2 ≤ 6 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 9
  - Class Estimated probability
    - 1: 0.667
    - 1: 0.333
Results on test data
- Covered examples: 3
  - Class Relative frequency
    - 1: 0.333
    - 1: 0.667

Rule 37
Predicted class is '1'
IF Wind Speed ≤ 11 AND Visibility (mi) > 3.5 AND Overcast Ceiling > 17.5 AND Overcast Ceiling ≤ 65 AND Scattered Ceiling_0 ≤ 32.5 AND Broken Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 6 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 19
  - Class Estimated probability
    - 1: 0.429
    - 1: 0.571
Results on test data
- Covered examples: 9
  - Class Relative frequency
    - 1: 0.333
    - 1: 0.667

Rule 38
Predicted class is '1'
IF Wind Speed ≤ 11 AND Visibility (mi) > 3.5 AND Overcast Ceiling > 17.5 AND Overcast Ceiling ≤ 65 AND Scattered Ceiling_0 ≤ 32.5 AND Broken Ceiling_0 > 32.5 AND Wind Speed_2 > 6 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 8
  - Class Estimated probability
    - 1: 0.750
    - 1: 0.250
Results on test data
- Covered examples: 5
  - Class Relative frequency
    - 1: 0.400
    - 1: 0.600

Rule 39
Predicted class is '1'
IF Wind Speed > 11 AND Visibility (mi) > 3.5 AND Overcast Ceiling > 17.5 AND Overcast Ceiling ≤ 65 AND Scattered Ceiling_0 ≤ 32.5 AND Broken Ceiling_0 ≤ 32.5 AND Wind Speed_2 > 6 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 16
  - Class Estimated probability
    - 1: 0.556
    - 1: 0.444
Results on test data
- Covered examples: 3
  - Class Relative frequency
    - 1: 0.333
    - 1: 0.667

Rule 40
Predicted class is '1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling > 65 AND Broken Ceiling ≤ 20 AND Scattered Ceiling_0 ≤ 32.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 15
  - Class Estimated probability
    - 1: 0.636
    - 1: 0.364
Results on test data
- Covered examples: 9
  - Class Relative frequency
    - 1: 0.556
    - 1: 0.444

Rule 41
Predicted class is '1'
IF Visibility (mi) > 3.5 AND Overcast Ceiling > 65 AND Broken Ceiling > 20 AND Scattered Ceiling_0 ≤ 32.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '1'

Results on training data
- Covered examples: 7
  - Class Estimated probability
    - 1: 0.250
    - 1: 0.750
Results on test data
- Covered examples: 6
  - Class Relative frequency
    - 1: 0.500
Rule 42
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 > 32.5 AND Broken Ceiling_0 ≤ 70 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
Covered examples: 147
Class Estimated probability
-1  0.333
  1  0.667
Covered examples: 73
Class Estimated probability
-1  0.077
  1  0.923

Rule 43
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Cross Winds ≤ 0.5 AND Scattered Ceiling_0 > 32.5 AND Broken Ceiling_0 > 70 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
Covered examples: 30
Class Estimated probability
-1  0.789
  1  0.211
Covered examples: 14
Class Estimated probability
-1  0.929
  1  0.071

Rule 44
Predicted class is '-1'
IF Wind Speed ≤ 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
Covered examples: 64
Class Estimated probability
-1  0.657
  1  0.343
Covered examples: 28
Class Estimated probability
-1  0.821
  1  0.179

Rule 45
Predicted class is '-1'
IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 ≤ 6.5 AND Visibility (mi)_2 > 5.5 THEN predicted = '-1'

Results on training data
Covered examples: 19
Class Estimated probability
-1  0.571
  1  0.429
Covered examples: 11
Class Estimated probability
-1  1.000
  1  0.000

Rule 46
Predicted class is '-1'
IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 6.5 AND Wind Speed_1 ≤ 9.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 AND Year ≤ 4.5 THEN predicted = '-1'

Results on training data
Covered examples: 31
Class Estimated probability
-1  0.533
  1  0.467
Covered examples: 12
Class Estimated probability
-1  0.417
  1  0.583

Rule 47
Predicted class is '-1'
IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 9.5 AND Wind Speed_1 ≤ 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 AND Year ≤ 4.5 THEN predicted = '-1'

Results on training data
Covered examples: 15
Class Estimated probability
-1  0.750
  1  0.250
Covered examples: 10
Class Estimated probability
-1  0.600
  1  0.400
### Rule 48

**Predicted class is -1**

IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 8.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 6.5 AND Wind Speed_1 ≤ 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 AND Year > 4.5 THEN predicted = -1

**Results on training data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>-1</td>
<td>0.900</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.100</td>
</tr>
</tbody>
</table>

**Results on test data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-1</td>
<td>0.571</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.429</td>
</tr>
</tbody>
</table>

### Rule 49

**Predicted class is -1**

IF Wind Speed > 3.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Wind Speed_0 > 8.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 6.5 AND Wind Speed_1 ≤ 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 AND Year > 4.5 THEN predicted = -1

**Results on training data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>-1</td>
<td>0.800</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.200</td>
</tr>
</tbody>
</table>

**Results on test data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-1</td>
<td>0.833</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.167</td>
</tr>
</tbody>
</table>

### Rule 50

**Predicted class is -1**

IF Wind Speed > 3.5 AND Wind Speed ≤ 7.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 THEN predicted = -1

**Results on training data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-1</td>
<td>0.250</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.750</td>
</tr>
</tbody>
</table>

**Results on test data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Rule 51

**Predicted class is 1**

IF Wind Speed > 7.5 AND Wind Speed ≤ 8.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 THEN predicted = 1

**Results on training data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>-1</td>
<td>0.833</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.167</td>
</tr>
</tbody>
</table>

**Results on test data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>-1</td>
<td>0.875</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.125</td>
</tr>
</tbody>
</table>

### Rule 52

**Predicted class is -1**

IF Wind Speed > 8.5 AND Visibility (mi) > 3.5 AND Cross Winds > 0.5 AND Scattered Ceiling_0 > 32.5 AND Wind Speed_1 > 10.5 AND Wind Speed_2 ≤ 20.5 AND Visibility (mi)_2 > 5.5 THEN predicted = -1

**Results on training data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>-1</td>
<td>0.556</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.444</td>
</tr>
</tbody>
</table>

**Results on test data**

<table>
<thead>
<tr>
<th>Covered examples</th>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-1</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>
1.5 AAR 48 at 1600

**Rule 1**
Predicted class is '1'
IF Cross Winds ≤ 0.5 AND Overcast Ceiling_0 ≤ 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'

**Results on training data**
- Covered examples: 13
  - Class: Estimated probability
    - 1: 0.125
    - 0: 0.875

**Results on test data**
- Covered examples: 6
  - Class: Relative frequency
    - 1: 0.333
    - 0: 0.667

**Rule 2**
Predicted class is '-1'
IF Cross Winds > 0.5 AND Overcast Ceiling_0 ≤ 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '-1'

**Results on training data**
- Covered examples: 12
  - Class: Estimated probability
    - 1: 0.625
    - 0: 0.375

**Results on test data**
- Covered examples: 4
  - Class: Relative frequency
    - 1: 0.500
    - 0: 0.500

**Rule 3**
Predicted class is '1'
IF Overcast Ceiling_0 > 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'

**Results on training data**
- Covered examples: 14
  - Class: Estimated probability
    - 1: 0.750
    - 0: 0.250

**Results on test data**
- Covered examples: 11
  - Class: Relative frequency
    - 1: 0.727
    - 0: 0.273

**Rule 4**
Predicted class is '1'
IF Overcast Ceiling ≤ 8.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 > 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'

**Results on training data**
- Covered examples: 57
  - Class: Estimated probability
    - 1: 0.421
    - 0: 0.579

**Results on test data**
- Covered examples: 19
  - Class: Relative frequency
    - 1: 0.421
    - 0: 0.579

**Rule 5**
Predicted class is '-1'
IF Overcast Ceiling > 0.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 > 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 AND Wind Speed_2 > 17 THEN predicted = '-1'

**Results on training data**
- Covered examples: 31
  - Class: Estimated probability
    - 1: 0.688
    - 0: 0.313

**Results on test data**
- Covered examples: 10
  - Class: Relative frequency
    - 1: 0.300
    - 0: 0.700

**Rule 6**
Predicted class is '1'
IF Overcast Ceiling > 0.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 > 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 AND Wind Speed_2 > 17 THEN predicted = '1'

**Results on training data**
- Covered examples: 11
  - Class: Estimated probability
    - 1: 0.125
    - 0: 0.875

**Results on test data**
- Covered examples: 5
  - Class: Relative frequency
    - 1: 0.400
    - 0: 0.600
Rule 7
Predicted class is '-1'
IF Overcast Ceiling_0 > 75 AND Wind Speed_1 > 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '-1'

Results on training data
- Covered examples: 11
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.525</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.375</td>
<td>1</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 2
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.500</td>
</tr>
<tr>
<td>1</td>
<td>0.500</td>
</tr>
</tbody>
</table>

Rule 8
Predicted class is '-1'
IF Cross Winds ≤ 0.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 > 9 THEN predicted = '-1'

Results on training data
- Covered examples: 12
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.500</td>
</tr>
<tr>
<td>1</td>
<td>0.500</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 9
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.778</td>
</tr>
<tr>
<td>1</td>
<td>0.222</td>
</tr>
</tbody>
</table>

Rule 9
Predicted class is '-1'
IF Cross Winds > 0.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 > 9 THEN predicted = '-1'

Results on training data
- Covered examples: 21
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.786</td>
</tr>
<tr>
<td>1</td>
<td>0.214</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 7
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.857</td>
</tr>
<tr>
<td>1</td>
<td>0.143</td>
</tr>
</tbody>
</table>

Rule 10
Predicted class is '-1'
IF Visibility (mi) ≤ 3.5 AND Broken Ceiling ≤ 1.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 57
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.899</td>
</tr>
<tr>
<td>1</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 35
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.771</td>
</tr>
<tr>
<td>1</td>
<td>0.229</td>
</tr>
</tbody>
</table>

Rule 11
Predicted class is '-1'
IF Visibility (mi) ≤ 3.5 AND Broken Ceiling > 1.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 17
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.875</td>
</tr>
<tr>
<td>1</td>
<td>0.125</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 5
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Rule 12
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Few Ceiling_1 ≤ 55 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 500
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.954</td>
</tr>
<tr>
<td>1</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Results on test data
- Covered examples: 200
  
<table>
<thead>
<tr>
<th>Class</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.955</td>
</tr>
<tr>
<td>1</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Rule 13
Predicted class is '-1'
IF Visibility (mi) > 3.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Few Ceiling_1 > 55 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 30
  - Class: Estimated probability
    - -1: 0.933
    - 1: 0.067

Results on test data
- Covered examples: 7
  - Class: Relative frequency
    - -1: 1.000
    - 1: 0.000

Rule 14
Predicted class is '-1'
IF Broken Ceiling_0 > 11.5 AND Wind Speed_1 ≤ 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 9
  - Class: Estimated probability
    - -1: 0.800
    - 1: 0.200

Results on test data
- Covered examples: 4
  - Class: Relative frequency
    - -1: 1.000
    - 1: 0.000

Rule 15
Predicted class is '-1'
IF Scattered Ceiling ≤ 37.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Scattered Ceiling_1 ≤ 17.5 AND Broken Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 67
  - Class: Estimated probability
    - -1: 0.854
    - 1: 0.146

Results on test data
- Covered examples: 30
  - Class: Relative frequency
    - -1: 1.000
    - 1: 0.167

Rule 16
Predicted class is '-1'
IF Scattered Ceiling ≤ 37.5 AND Cross Winds ≤ 0.5 AND Wind Speed_0 ≤ 24.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Broken Ceiling_1 > 37.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 THEN predicted = '-1'

Results on training data
- Covered examples: 38
  - Class: Estimated probability
    - -1: 0.958
    - 1: 0.042

Results on test data
- Covered examples: 16
  - Class: Relative frequency
    - -1: 1.000
    - 1: 0.000

Rule 17
Predicted class is '-1'
IF Scattered Ceiling ≤ 37.5 AND Cross Winds > 0.5 AND Wind Speed_0 ≤ 24.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Broken Ceiling_1 > 37.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 THEN predicted = '-1'

Results on training data
- Covered examples: 69
  - Class: Estimated probability
    - -1: 0.925
    - 1: 0.175

Results on test data
- Covered examples: 31
  - Class: Relative frequency
    - -1: 0.806
    - 1: 0.194

Rule 18
Predicted class is '-1'
IF Scattered Ceiling ≤ 37.5 AND Wind Speed_0 > 24.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Broken Ceiling_1 > 37.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 ≤ 55 THEN predicted = '-1'

Results on training data
- Covered examples: 11
  - Class: Estimated probability
    - -1: 0.833
    - 1: 0.167

Results on test data
- Covered examples: 6
  - Class: Relative frequency
    - -1: 0.833
    - 1: 0.167
<table>
<thead>
<tr>
<th>Rule</th>
<th>Predicted class is '1'</th>
<th>Predicted class is '1'</th>
<th>Predicted class is '1'</th>
<th>Predicted class is '1'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 25</td>
<td>IF Rain (Y/N) ≤ 0.5 AND Visibility (m)₁ ≤ 5.5 AND Thunder storms (Y/N)₁ ≤ 0.5 AND Mist (Y/N)₁ ≤ 0.5 AND Rain (Y/N)₂ &gt; 0.5 AND Broken Ceiling₂ ≤ 165 AND Year ≤ 5.5 THEN predicted = '1'</td>
<td>IF Rain (Y/N) &gt; 0.5 AND Visibility (m)₁ ≤ 5.5 AND Thunder storms (Y/N)₁ ≤ 0.5 AND Mist (Y/N)₁ ≤ 0.5 AND Rain (Y/N)₂ &gt; 0.5 AND Broken Ceiling₂ ≤ 165 AND Year ≤ 5.5 THEN predicted = '1'</td>
<td>IF Visibility (m)₁ &gt; 5.5 AND Thunder storms (Y/N)₁ &gt; 0.5 AND Mist (Y/N)₁ &gt; 0.5 AND Rain (Y/N)₂ &gt; 0.5 AND Broken Ceiling₂ ≤ 165 AND Year ≤ 5.5 THEN predicted = '1'</td>
<td>IF Visibility (m)₁ &gt; 5.5 AND Rain (Y/N)₂ &gt; 0.5 AND Broken Ceiling₂ ≤ 165 AND Year ≤ 5.5 THEN predicted = '1'</td>
</tr>
<tr>
<td>Results on training data</td>
<td>Results on test data</td>
<td>Results on training data</td>
<td>Results on test data</td>
<td>Results on training data</td>
</tr>
<tr>
<td>Covered examples: 30</td>
<td>Covered examples: 50</td>
<td>Covered examples: 17</td>
<td>Covered examples: 17</td>
<td>Covered examples: 13</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
</tr>
<tr>
<td>-1</td>
<td>0.929</td>
<td>-1</td>
<td>0.600</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>0.071</td>
<td>1</td>
<td>0.400</td>
<td>1</td>
</tr>
<tr>
<td>Covered examples: 31</td>
<td>Covered examples: 8</td>
<td>Covered examples: 3</td>
<td>Covered examples: 7</td>
<td>Covered examples: 9</td>
</tr>
<tr>
<td>Class Relative frequency</td>
<td>Class Relative frequency</td>
<td>Class Relative frequency</td>
<td>Class Relative frequency</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1</td>
<td>0.903</td>
<td>-1</td>
<td>0.625</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>0.097</td>
<td>1</td>
<td>0.375</td>
<td>1</td>
</tr>
<tr>
<td>Covered examples: 9</td>
<td>Covered examples: 19</td>
<td>Covered examples: 4</td>
<td>Covered examples: 17</td>
<td>Covered examples: 20</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
<td>Class Estimated probability</td>
</tr>
<tr>
<td>-1</td>
<td>0.643</td>
<td>-1</td>
<td>0.538</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>0.357</td>
<td>1</td>
<td>0.462</td>
<td>1</td>
</tr>
</tbody>
</table>
### I.6 AAR 52 at 1600

**Rule 1**
Predicted class is '1'.

If Cross Winds ≤ 0.5 AND Overcast Ceiling_0 ≤ 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 13</td>
<td>Covered examples: 6</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.125</td>
<td>-1: 0.333</td>
</tr>
<tr>
<td>1: 0.875</td>
<td>1: 0.667</td>
</tr>
</tbody>
</table>

**Rule 2**
Predicted class is '-1'.

If Cross Winds > 0.5 AND Overcast Ceiling_0 ≤ 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '-1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 12</td>
<td>Covered examples: 4</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.625</td>
<td>-1: 0.500</td>
</tr>
<tr>
<td>1: 0.375</td>
<td>1: 0.500</td>
</tr>
</tbody>
</table>

**Rule 3**
Predicted class is '1'.

If Overcast Ceiling_0 > 12.5 AND Wind Speed_1 ≤ 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 14</td>
<td>Covered examples: 11</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.750</td>
<td>-1: 0.727</td>
</tr>
<tr>
<td>1: 0.250</td>
<td>1: 0.273</td>
</tr>
</tbody>
</table>

**Rule 4**
Predicted class is '1'.

If Overcast Ceiling ≤ 8.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 = 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 THEN predicted = '1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 57</td>
<td>Covered examples: 19</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.421</td>
<td>-1: 0.421</td>
</tr>
<tr>
<td>1: 0.579</td>
<td>1: 0.579</td>
</tr>
</tbody>
</table>

**Rule 5**
Predicted class is '-1'.

If Overcast Ceiling > 8.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 = 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 AND Wind Speed_2 > 17 THEN predicted = '-1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 31</td>
<td>Covered examples: 10</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.688</td>
<td>-1: 0.300</td>
</tr>
<tr>
<td>1: 0.313</td>
<td>1: 0.700</td>
</tr>
</tbody>
</table>

**Rule 6**
Predicted class is '1'.

If Overcast Ceiling > 8.5 AND Overcast Ceiling_0 ≤ 75 AND Wind Speed_1 = 7.5 AND Visibility (mi)_1 ≤ 5.5 AND Broken Ceiling_1 ≤ 9 AND Wind Speed_2 > 17 THEN predicted = '1'.

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 11</td>
<td>Covered examples: 5</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Estimated probability</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>-1: 0.125</td>
<td>-1: 0.400</td>
</tr>
<tr>
<td>1: 0.875</td>
<td>1: 0.600</td>
</tr>
</tbody>
</table>
Rule 7
Predicted class is '-1'

IF Overcast Ceiling_0 > 0.5 AND Wind Speed_1 > 7.5 AND Visibility (mi)_1 ≤ 0.5 AND Broken Ceiling_1 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 11
  - Class Estimated probability
    - 1 0.375
    - -1 0.625

Results on test data
- Covered examples: 2
  - Class Relative frequency
    - 1 0.500
    - -1 0.500

Rule 8
Predicted class is '-1'

IF Cross Winds ≤ 0.5 AND Visibility (mi)_1 ≤ 0.5 AND Broken Ceiling_1 > 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 12
  - Class Estimated probability
    - 1 0.500
    - -1 0.500

Results on test data
- Covered examples: 9
  - Class Relative frequency
    - 1 0.222
    - -1 0.778

Rule 9
Predicted class is '-1'

IF Cross Winds > 0.5 AND Visibility (mi)_1 ≤ 0.5 AND Broken Ceiling_1 > 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 21
  - Class Estimated probability
    - 1 0.214
    - -1 0.786

Results on test data
- Covered examples: 7
  - Class Relative frequency
    - 1 0.143
    - -1 0.857

Rule 10
Predicted class is '-1'

IF Visibility (mi)_1 ≤ 0.5 AND Broken Ceiling_1 ≤ 0.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 0.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 57
  - Class Estimated probability
    - 1 0.111
    - -1 0.889

Results on test data
- Covered examples: 35
  - Class Relative frequency
    - 1 0.229
    - -1 0.771

Rule 11
Predicted class is '-1'

IF Visibility (mi)_1 ≤ 0.5 AND Broken Ceiling_1 > 0.5 AND Broken Ceiling_0 ≤ 11.5 AND Visibility (mi)_1 > 0.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 17
  - Class Estimated probability
    - 1 0.125
    - -1 0.875

Results on test data
- Covered examples: 5
  - Class Relative frequency
    - 1 0.000
    - -1 1.000

Rule 12
Predicted class is '-1'

IF Visibility (mi)_1 > 0.5 AND Broken Ceiling_0 ≤ 0.5 AND Visibility (mi)_1 > 0.5 AND Overcast Ceiling_1 ≤ 17.5 AND Few Ceiling_1 ≤ 0.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 900
  - Class Estimated probability
    - 1 0.046
    - -1 0.954

Results on test data
- Covered examples: 200
  - Class Relative frequency
    - 1 0.045
    - -1 0.955
Rule 19
Predicted class is '1'.
IF Scattered Ceiling ≤ 37.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Broken Ceiling_1 > 37.5 AND Rain (Y/N)_2 ≤ 0.5 AND Scattered Ceiling_2 > 55 THEN predicted = '1'.

Results on training data
Covered examples: 19
Class Estimated probability
-1 0.875
1 0.125

Results on test data
Covered examples: 8
Class Relative frequency
-1 1.000
1 0.000

Rule 20
Predicted class is '1'.
IF Scattered Ceiling > 37.5 AND Broken Ceiling_0 > 11.5 AND Wind Speed_1 > 3.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 ≤ 0.5 THEN predicted = '1'.

Results on training data
Covered examples: 64
Class Estimated probability
-1 0.941
1 0.059

Results on test data
Covered examples: 25
Class Relative frequency
-1 1.000
1 0.000

Rule 21
Predicted class is '1'.
IF Visibility (mi)_1 = 5.5 AND Overcast Ceiling_1 > 17.5 AND Rain (Y/N)_2 ≤ 0.5 AND Overcast Ceiling_2 ≤ 3.5 THEN predicted = '1'.

Results on training data
Covered examples: 19
Class Estimated probability
-1 0.667
1 0.333

Results on test data
Covered examples: 5
Class Relative frequency
-1 1.000
1 0.000

Rule 22
Predicted class is '1'.
IF Overcast Ceiling_0 ≤ 10 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 > 17.5 AND Rain (Y/N)_2 ≤ 0.5 AND Overcast Ceiling_2 > 3.5 THEN predicted = '1'.

Results on training data
Covered examples: 18
Class Estimated probability
-1 0.667
1 0.333

Results on test data
Covered examples: 10
Class Relative frequency
-1 0.800
1 0.200

Rule 23
Predicted class is '1'.
IF Overcast Ceiling_0 > 10 AND Cross Winds_0 ≤ 0.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 > 17.5 AND Rain (Y/N)_2 ≤ 0.5 AND Overcast Ceiling_2 > 3.5 THEN predicted = '1'.

Results on training data
Covered examples: 17
Class Estimated probability
-1 0.900
1 0.100

Results on test data
Covered examples: 13
Class Relative frequency
-1 1.000
1 0.000

Rule 24
Predicted class is '1'.
IF Overcast Ceiling_0 > 10 AND Cross Winds_0 > 0.5 AND Visibility (mi)_1 > 5.5 AND Overcast Ceiling_1 > 17.5 AND Rain (Y/N)_2 ≤ 0.5 AND Overcast Ceiling_2 > 3.5 THEN predicted = '1'.

Results on training data
Covered examples: 23
Class Estimated probability
-1 0.786
1 0.214

Results on test data
Covered examples: 15
Class Relative frequency
-1 0.733
1 0.267
Rule 25
Predicted class is "-1"
IF Rain (Y/N) ≤ 0.5 AND Visibility (m) ≤ 5.5 AND Thunder storms (Y/N) ≤ 0.5 AND Mist (Y/N) ≤ 0.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling ≤ 165 AND Year ≤ 5.5 THEN predicted = "-1"

Results on training data
Covered examples: 50
-1 0.929
1 0.071

Results on test data
Covered examples: 31
-1 0.903
1 0.097

Rule 26
Predicted class is "-1"
IF Rain (Y/N) > 0.5 AND Visibility (m) ≤ 5.5 AND Thunder storms (Y/N) ≤ 0.5 AND Mist (Y/N) ≤ 0.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling ≤ 165 AND Year ≤ 5.5 THEN predicted = "-1"

Results on training data
Covered examples: 17
-1 0.600
1 0.400

Results on test data
Covered examples: 8
-1 0.625
1 0.375

Rule 27
Predicted class is "-1"
IF Visibility (m) ≥ 5.5 AND Thunder storms (Y/N) ≥ 0.5 AND Mist (Y/N) ≤ 0.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling ≤ 165 AND Year ≤ 5.5 THEN predicted = "-1"

Results on training data
Covered examples: 17
-1 0.778
1 0.222

Results on test data
Covered examples: 3
-1 0.667
1 0.333

Rule 28
Predicted class is "-1"
IF Visibility (m) ≥ 5.5 AND Mist (Y/N) ≥ 0.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling ≤ 165 AND Year ≤ 5.5 THEN predicted = "-1"

Results on training data
Covered examples: 13
-1 0.714
1 0.286

Results on test data
Covered examples: 7
-1 0.857
1 0.143

Rule 29
Predicted class is "-1"
IF Visibility (m) ≥ 5.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling ≤ 165 AND Year > 5.5 THEN predicted = "-1"

Results on training data
Covered examples: 22
-1 0.538
1 0.462

Results on test data
Covered examples: 6
-1 0.667
1 0.333

Rule 30
Predicted class is "-1"
IF Visibility (m) ≥ 5.5 AND Rain (Y/N) ≥ 0.5 AND Broken Ceiling > 165 THEN predicted = "-1"

Results on training data
Covered examples: 20
-1 0.643
1 0.357

Results on test data
Covered examples: 9
-1 0.778
1 0.222
I.7 AAR 48 at 1800

Rule 1
Predicted class is ‘-1’
IF Overcast Ceiling\_0 ≤ 1.5 AND Mist (Y/N)\_1 ≤ 0.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 921
  - Class: Estimated probability
    - -1: 0.912
    - 1: 0.088

Results on test data
- Covered examples: 413
  - Class: Relative frequency
    - -1: 0.918
    - 1: 0.082

Rule 2
Predicted class is ‘-1’
IF Rain (Y/N) ≤ 0.5 AND Overcast Ceiling\_0 > 1.5 AND Mist (Y/N)\_1 ≤ 0.5 AND Snow (Y/N)\_2 ≤ 0.5 THEN predicted = ‘-1’

Results on training data
- Covered examples: 130
  - Class: Estimated probability
    - -1: 0.809
    - 1: 0.191

Results on test data
- Covered examples: 49
  - Class: Relative frequency
    - -1: 0.837
    - 1: 0.163

Rule 3
Predicted class is ‘1’
IF Wind Speed ≤ 7.5 AND Rain (Y/N) > 0.5 AND Showers (Y/N) ≤ 0.5 AND Overcast Ceiling\_0 > 1.5 AND Mist (Y/N)\_1 ≤ 0.5 AND Snow (Y/N)\_2 ≤ 0.5 THEN predicted = ‘1’

Results on training data
- Covered examples: 15
  - Class: Estimated probability
    - -1: 0.333
    - 1: 0.667

Results on test data
- Covered examples: 4
  - Class: Relative frequency
    - -1: 0.500
    - 1: 0.500

Rule 4
Predicted class is ‘-1’
IF Wind Speed > 7.5 AND Rain (Y/N) > 0.5 AND Showers (Y/N) ≤ 0.5 AND Overcast Ceiling\_0 > 1.5 AND Mist (Y/N)\_1 ≤ 0.5 AND Snow (Y/N)\_2 ≤ 0.5 AND Broken Ceiling\_2 ≤ 6 THEN predicted = ‘-1’

Results on training data
- Covered examples: 13
  - Class: Estimated probability
    - -1: 0.857
    - 1: 0.143

Results on test data
- Covered examples: 5
  - Class: Relative frequency
    - -1: 0.800
    - 1: 0.200

Rule 5
Predicted class is ‘-1’
IF Wind Speed > 7.5 AND Rain (Y/N) > 0.5 AND Showers (Y/N) ≤ 0.5 AND Overcast Ceiling\_0 > 1.5 AND Mist (Y/N)\_1 ≤ 0.5 AND Snow (Y/N)\_2 ≤ 0.5 AND Broken Ceiling\_2 > 6 THEN predicted = ‘-1’

Results on training data
- Covered examples: 12
  - Class: Estimated probability
    - -1: 0.625
    - 1: 0.375

Results on test data
- Covered examples: 3
  - Class: Relative frequency
    - -1: 1.000
    - 1: 0.000

Rule 6
Predicted class is ‘1’
IF Rain (Y/N) > 0.5 AND Showers (Y/N) > 0.5 AND Overcast Ceiling\_0 > 1.5 AND Mist (Y/N)\_1 ≤ 0.5 AND Snow (Y/N)\_2 ≤ 0.5 THEN predicted = ‘1’

Results on training data
- Covered examples: 15
  - Class: Estimated probability
    - -1: 0.545
    - 1: 0.455

Results on test data
- Covered examples: 9
  - Class: Relative frequency
    - -1: 0.889
    - 1: 0.111
Rule 7
Predicted class is '-1'
IF Overcast Ceiling_0 > 1.5 AND Mist (Y/N)_1 ≤ 0.5 AND Snow (Y/N)_2 > 0.5 THEN predicted = '-1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 17</td>
<td>Covered examples: 4</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.667</td>
<td>-1 0.250</td>
</tr>
<tr>
<td>1 0.333</td>
<td>1 0.750</td>
</tr>
</tbody>
</table>

Rule 8
Predicted class is '1'
IF Wind Speed ≤ 5.5 AND Scattered Ceiling_0 ≤ 2.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 ≤ 30 AND Overcast Ceiling_2 ≤ 6.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 16</td>
<td>Covered examples: 5</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.375</td>
<td>-1 0.200</td>
</tr>
<tr>
<td>1 0.625</td>
<td>1 0.800</td>
</tr>
</tbody>
</table>

Rule 9
Predicted class is '1'
IF Wind Speed > 5.5 AND Wind Speed ≤ 9.5 AND Scattered Ceiling_0 ≤ 2.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 ≤ 30 AND Overcast Ceiling_2 ≤ 6.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 23</td>
<td>Covered examples: 8</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.600</td>
<td>-1 0.500</td>
</tr>
<tr>
<td>1 0.400</td>
<td>1 0.500</td>
</tr>
</tbody>
</table>

Rule 10
Predicted class is '1'
IF Wind Speed ≤ 9.5 AND Scattered Ceiling_0 > 2.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 ≤ 30 AND Overcast Ceiling_2 ≤ 6.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 16</td>
<td>Covered examples: 4</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.818</td>
<td>-1 0.500</td>
</tr>
<tr>
<td>1 0.182</td>
<td>1 0.500</td>
</tr>
</tbody>
</table>

Rule 11
Predicted class is '1'
IF Wind Speed ≤ 9.5 AND Rain (Y/N) ≤ 0.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 ≤ 13.5 AND Overcast Ceiling_2 > 6.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 13</td>
<td>Covered examples: 4</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.667</td>
<td>-1 0.500</td>
</tr>
<tr>
<td>1 0.333</td>
<td>1 0.500</td>
</tr>
</tbody>
</table>

Rule 12
Predicted class is '1'
IF Wind Speed ≤ 9.5 AND Rain (Y/N) > 0.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 ≤ 13.5 AND Overcast Ceiling_2 > 6.5 THEN predicted = '1'

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 11</td>
<td>Covered examples: 6</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.250</td>
<td>-1 0.500</td>
</tr>
<tr>
<td>1 0.750</td>
<td>1 0.500</td>
</tr>
</tbody>
</table>
Rule 13
Predicted class is '-1'
IF Wind Speed ≤ 9.5 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 > 13.5 AND Overcast Ceiling_2 > 6.5 THEN predicted = '-1'

Results on training data
Covered examples: 19
Class Estimated probability
-1 0.583
1 0.417

Results on test data
Covered examples: 10
Class Relative frequency
-1 0.600
1 0.400

Rule 14
Predicted class is '-1'
IF Wind Speed ≤ 9.5 AND Broken Ceiling ≤ 2 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 > 30 THEN predicted = '-1'

Results on training data
Covered examples: 16
Class Estimated probability
-1 0.700
1 0.300

Results on test data
Covered examples: 5
Class Relative frequency
-1 0.600
1 0.400

Rule 15
Predicted class is '-1'
IF Wind Speed ≤ 9.5 AND Broken Ceiling > 2 AND Mist (Y/N)_1 > 0.5 AND Overcast Ceiling_1 > 30 THEN predicted = '-1'

Results on training data
Covered examples: 8
Class Estimated probability
-1 0.800
1 0.200

Results on test data
Covered examples: 4
Class Relative frequency
-1 0.250
1 0.750

Rule 16
Predicted class is '1'
IF Wind Speed > 9.5 AND Visibility (mi)_0 ≤ 3.5 AND Mist (Y/N)_1 > 0.5 THEN predicted = '1'

Results on training data
Covered examples: 19
Class Estimated probability
-1 0.250
1 0.750

Results on test data
Covered examples: 11
Class Relative frequency
-1 0.273
1 0.727

Rule 17
Predicted class is '-1'
IF Wind Speed > 9.5 AND Visibility (mi)_0 > 3.5 AND Mist (Y/N)_1 > 0.5 THEN predicted = '-1'

Results on training data
Covered examples: 15
Class Estimated probability
-1 0.556
1 0.444

Results on test data
Covered examples: 3
Class Relative frequency
-1 0.667
1 0.333
1.8 AAR 52 at 1800

### Rule 1

Predicted class is ‘-1’

- **If Wind Speed \(\leq 7.5\) AND Visibility (mi) \(0 \leq 5.5\) AND Wind Speed \(_1\) \(\leq 10.5\) AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) THEN predicted = ’-1’**

#### Results on training data

- Covered examples: 35
- Class Estimated probability
  - 1: 0.657
  - 0: 0.333

#### Results on test data

- Covered examples: 23
- Class Estimated probability
  - 1: 0.522
  - 0: 0.478

### Rule 2

Predicted class is ‘1’

- **If Wind Speed \(\leq 7.5\) AND Visibility (mi) \(0 \leq 5.5\) AND Wind Speed \(_1\) > 10.5 AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) THEN predicted = ’1’**

#### Results on training data

- Covered examples: 11
- Class Estimated probability
  - 1: 0.750
  - 0: 0.250

#### Results on test data

- Covered examples: 5
- Class Estimated probability
  - 1: 0.200
  - 0: 0.800

### Rule 3

Predicted class is ‘-1’

- **If Wind Speed > 7.5 AND Visibility (mi) \(0 \geq 5.5\) AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) THEN predicted = ’-1’**

#### Results on training data

- Covered examples: 19
- Class Estimated probability
  - 1: 0.556
  - 0: 0.444

#### Results on test data

- Covered examples: 9
- Class Estimated probability
  - 1: 0.444
  - 0: 0.556

### Rule 4

Predicted class is ‘-1’

- **If Broken Ceiling \(\leq 22.5\) AND Cross Winds \(\leq 0.5\) AND Visibility (mi) \(0 > 5.5\) AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) THEN predicted = ’-1’**

#### Results on training data

- Covered examples: 259
- Class Estimated probability
  - 1: 0.109
  - 0: 0.891

#### Results on test data

- Covered examples: 110
- Class Estimated probability
  - 1: 0.100
  - 0: 0.900

### Rule 5

Predicted class is ‘-1’

- **If Scattered Ceiling \(\leq 20\) AND Broken Ceiling \(> 22.5\) AND Broken Ceiling \(_1\) \(\leq 45\) AND Cross Winds \(\leq 0.5\) AND Visibility (mi) \(0 > 5.5\) AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) AND Broken Ceiling \(_2\) \(\leq 27.5\) THEN predicted = ’-1’**

#### Results on training data

- Covered examples: 11
- Class Estimated probability
  - 1: 0.167
  - 0: 0.833

#### Results on test data

- Covered examples: 10
- Class Estimated probability
  - 1: 0.200
  - 0: 0.800

### Rule 6

Predicted class is ‘-1’

- **If Scattered Ceiling \(\leq 20\) AND Broken Ceiling \(> 45\) AND Cross Winds \(\leq 0.5\) AND Visibility (mi) \(0 > 5.5\) AND Overcast Ceiling \(_1\) \(\leq 2\) AND Rain (Y/N) \(_2\) \(\leq 0.5\) AND Broken Ceiling \(_2\) \(\leq 27.5\) THEN predicted = ’-1’**

#### Results on training data

- Covered examples: 22
- Class Estimated probability
  - 1: 0.308
  - 0: 0.692

#### Results on test data

- Covered examples: 12
- Class Estimated probability
  - 1: 0.000
  - 0: 1.000
Rule 18
Predicted class is '-1'
IF Wind Speed ≤ 9.5 AND Scattered Ceiling ≤ 4 AND Cross Winds > 0.5 AND Wind Speed > 4.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling ≥ 2 AND Broken Ceiling < 12.5 AND Few Ceiling ≤ 20 AND Wind Speed ≥ 11 AND Rain (Y/N) ≥ 0.5 AND Scattered Ceiling ≥ 20 THEN predicted = '-1'

Results on training data
- Covered examples: 65
  - Class Estimated probability
    - -1: 0.632
    - 1: 0.368

Results on test data
- Covered examples: 28
  - Class Relative frequency
    - -1: 0.714
    - 1: 0.286

Rule 19
Predicted class is '-1'
IF Wind Speed ≤ 9.5 AND Scattered Ceiling ≤ 4 AND Cross Winds > 0.5 AND Wind Speed > 4.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling ≥ 2 AND Broken Ceiling < 12.5 AND Few Ceiling ≤ 20 AND Wind Speed ≥ 11 AND Rain (Y/N) ≥ 0.5 AND Scattered Ceiling ≥ 20 THEN predicted = '-1'

Results on training data
- Covered examples: 16
  - Class Estimated probability
    - -1: 0.599
    - 1: 0.111

Results on test data
- Covered examples: 5
  - Class Relative frequency
    - -1: 0.800
    - 1: 0.200

Rule 20
Predicted class is '-1'
IF Wind Speed > 9.5 AND Scattered Ceiling ≤ 4 AND Cross Winds > 0.5 AND Wind Speed > 4.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling ≥ 2 AND Broken Ceiling < 12.5 AND Few Ceiling ≤ 20 AND Rain (Y/N) ≥ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 32
  - Class Estimated probability
    - -1: 0.800
    - 1: 0.200

Results on test data
- Covered examples: 17
  - Class Relative frequency
    - -1: 0.647
    - 1: 0.353

Rule 21
Predicted class is '-1'
IF Scattered Ceiling ≤ 4 AND Cross Winds > 0.5 AND Wind Speed > 4.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling ≥ 2 AND Broken Ceiling < 12.5 AND Broken Ceiling ≤ 70 AND Few Ceiling ≤ 20 AND Rain (Y/N) ≥ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 40
  - Class Estimated probability
    - -1: 0.500
    - 1: 0.500

Results on test data
- Covered examples: 20
  - Class Relative frequency
    - -1: 0.400
    - 1: 0.600

Rule 22
Predicted class is '-1'
IF Scattered Ceiling ≤ 4 AND Cross Winds > 0.5 AND Wind Speed > 4.5 AND Visibility (mi) > 5.5 AND Overcast Ceiling ≥ 2 AND Broken Ceiling ≤ 70 AND Few Ceiling ≥ 20 AND Rain (Y/N) ≥ 0.5 THEN predicted = '-1'

Results on training data
- Covered examples: 41
  - Class Estimated probability
    - -1: 0.773
    - 1: 0.227

Results on test data
- Covered examples: 17
  - Class Relative frequency
    - -1: 0.706
    - 1: 0.294
### Rule 29
Predicted class is ‘-1’
IF Broken Ceiling_0 ≤ 110 AND Overcast Ceiling_1 ≤ 2 AND Broken Ceiling_1 ≤ 17.5 AND Rain (Y/N)_2 > 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = ‘-1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 14</td>
<td>Covered examples: 2</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.500</td>
<td>-1 0.000</td>
</tr>
<tr>
<td>1 0.500</td>
<td>1 1.000</td>
</tr>
</tbody>
</table>

### Rule 30
Predicted class is ‘1’
IF Broken Ceiling_0 ≤ 27.5 AND Overcast Ceiling_1 ≤ 2 AND Broken Ceiling_1 > 17.5 AND Rain (Y/N)_2 > 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = ‘1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 19</td>
<td>Covered examples: 5</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.286</td>
<td>-1 0.400</td>
</tr>
<tr>
<td>1 0.714</td>
<td>1 0.400</td>
</tr>
</tbody>
</table>

### Rule 31
Predicted class is ‘-1’
IF Broken Ceiling_0 > 27.5 AND Broken Ceiling_0 ≤ 110 AND Overcast Ceiling_1 ≤ 2 AND Broken Ceiling_1 > 17.5 AND Rain (Y/N)_2 > 0.5 AND Showers (Y/N)_2 > 0.5 THEN predicted = ‘-1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 16</td>
<td>Covered examples: 7</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.667</td>
<td>-1 0.429</td>
</tr>
<tr>
<td>1 0.333</td>
<td>1 0.571</td>
</tr>
</tbody>
</table>

### Rule 32
Predicted class is ‘-1’
IF Broken Ceiling_0 > 110 AND Overcast Ceiling_1 ≤ 2 AND Cross Winds_1 ≤ 0.5 AND Rain (Y/N)_2 > 0.5 THEN predicted = ‘-1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 9</td>
<td>Covered examples: 5</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.667</td>
<td>-1 0.800</td>
</tr>
<tr>
<td>1 0.333</td>
<td>1 0.200</td>
</tr>
</tbody>
</table>

### Rule 33
Predicted class is ‘1’
IF Broken Ceiling_0 > 110 AND Overcast Ceiling_1 ≤ 2 AND Cross Winds_1 > 0.5 AND Rain (Y/N)_2 > 0.5 THEN predicted = ‘1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 10</td>
<td>Covered examples: 3</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.333</td>
<td>-1 0.333</td>
</tr>
<tr>
<td>1 0.667</td>
<td>1 0.667</td>
</tr>
</tbody>
</table>

### Rule 34
Predicted class is ‘1’
IF Overcast Ceiling_0 ≤ 32.5 AND Mist (Y/N)_1 ≤ 0.5 AND Overcast Ceiling_1 > 2 AND Cross Winds_1 ≤ 0.5 AND Broken Ceiling_2 ≤ 37.5 THEN predicted = ‘1’

<table>
<thead>
<tr>
<th>Results on training data</th>
<th>Results on test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered examples: 39</td>
<td>Covered examples: 13</td>
</tr>
<tr>
<td>Class Estimated probability</td>
<td>Class Relative frequency</td>
</tr>
<tr>
<td>-1 0.160</td>
<td>-1 0.385</td>
</tr>
<tr>
<td>1 0.840</td>
<td>1 0.615</td>
</tr>
<tr>
<td>Rule</td>
<td>Predicted class</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>35</td>
<td>'1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>'-1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>'1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>'1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>'-1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>'1'</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rule 41
Predicted class is '1'
IF Wind Speed > 5.5 AND Mist (Y/N) > 0.5 AND Overcast Ceiling > 2 THEN predicted = '1'

Results on training data
Covered examples: 76
Class  Estimated probability
  -1  0.175
    1  0.825

Results on test data
Covered examples: 25
Class  Relative frequency
  -1  0.120
    1  0.880
Appendix J: Prediction Vectors

J.1 Philadelphia Prediction Vector

\( w\)-vector at divider point 48 for time period 0800

\[
w = \begin{bmatrix} 0.027, -0.262, 1.241, 1.100, -0.899, 0.963, 0.353, -0.428, 1.482, -0.004, \\
0.000, 0.000, 0.000, -0.035, -0.020, -0.260, -0.777, -1.278, 0.702, 2.097, \\
-1.022, 0.037, -0.867, -0.001, -0.001, -0.001, 0.000, 0.078, 0.001, 0.049, \\
0.448, 0.627, -0.359, -0.052, 0.037, -0.104, -0.518, 0.000, -0.001, 0.000, \\
-0.004, -0.093, 0.013, -0.088, -0.144, 0.273, 0.156, -0.169, 0.000, -0.343, \\
0.500, 0.002, 0.000, 0.001, -0.004, 0.017, 0.128 \end{bmatrix}
\]

\( b = 1.538 \)

\( w\)-vector at divider point 52 for time period 0800

\[
w = \begin{bmatrix} 0.012, -0.436, 0.849, 0.332, -0.220, 0.000, -0.075, -0.317, 0.214, 0.001, \\
0.000, 0.000, 0.000, 0.024, -0.012, -0.063, 0.418, -0.295, 0.068, 0.455, \\
-2.016, 0.129, 0.000, 0.001, -0.001, -0.001, 0.000, 0.021, -0.004, -0.182, \\
-0.582, 0.541, 0.494, 1.026, 0.516, 0.767, -0.003, -0.001, 0.000, 0.001, \\
-0.002, 0.072, 0.012, -0.003, 0.802, 0.466, -0.775, -0.938, 0.000, 0.098, \\
-0.002, 0.001, -0.001, 0.000, 0.000, -0.070, 0.030 \end{bmatrix}
\]

\( b = 2.901 \)
\( w \)-vector at divider point 48 for time period 1200

\[
\begin{align*}
w &= [0.000, 0.000, 2.000, 0.000, -2.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] \\
b &= -1.000
\end{align*}
\]

\( w \)-vector at divider point 52 for time period 1200

\[
\begin{align*}
w &= [0.000, 0.000, 0.000, 0.954, 0.101, -0.101, 0.151, 1.125, 0.000, 0.690, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.111, 0.265, 0.288, 0.563, 0.151, -0.622, 0.603, -0.159, 0.002, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.160, 0.271, 0.280, -0.431, -0.271, -0.183, 0.026, -0.390, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] \\
b &= 0.457
\end{align*}
\]
\( w \)-vector at divider point 48 for time period 1600

\[
\begin{align*}
w &= [0.003, 0.013, 0.644, 0.569, -0.648, -0.452, 0.097, 0.127, 0.837, 0.000, \\
0.000, 0.001, 0.000, -0.004, 0.000, -0.212, 0.090, 0.075, 0.045, 0.733, \\
-0.726, -0.441, 0.864, -0.003, -0.001, -0.002, -0.001, -0.173, -0.008, 0.067, \\
0.338, 0.978, -0.386, -0.360, 1.098, 0.594, -0.362, -0.001, -0.001, 0.001, \\
0.000, 0.230, 0.017, -0.178, -0.253, 0.289, 0.394, 0.319, 0.000, -0.422, \\
-1.225, 0.000, 0.000, -0.001, -0.003, 0.000, 0.123]
\end{align*}
\]

\[
b = 0.004
\]

\( w \)-vector at divider point 52 for time period 1600

\[
\begin{align*}
w &= [0.000, 0.000, 0.810, 0.810, -0.810, -0.019, 1.264, 0.000, 0.194, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.032, \\
-1.100, 0.000, 0.194, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.032, \\
0.019, -0.238, 0.135, 0.000, 1.441, 1.010, -1.070, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, -0.141, 0.772, 1.576, 0.000, 0.000, 0.000, 0.064, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] \\
\end{align*}
\]

\[
b = 0.852
\]
\( w \)-vector at divider point 48 for time period 1800

\[
\begin{align*}
w & = [0.001, 0.026, 0.777, 0.027, -0.761, -0.784, 0.027, 0.091, -0.779, 0.000, \\
& \quad 0.000, 0.001, -0.001, -0.006, -0.003, -0.200, -0.023, 0.326, 0.029, 0.646, \\
& \quad -0.623, -0.406, 0.524, -0.001, -0.001, -0.002, -0.002, -0.178, 0.000, 0.072, \\
& \quad 0.647, 1.074, -0.658, -0.742, 0.914, 0.663, -0.369, 0.000, 0.000, 0.001, \\
& \quad 0.000, 0.243, 0.009, -0.155, -0.282, 0.338, 0.328, 0.588, 0.000, -0.273, \\
& \quad -1.158, -0.001, 0.000, -0.001, -0.004, -0.001, 0.143]
\end{align*}
\]

\( b = -0.389 \)

\( w \)-vector at divider point 52 for time period 1800

\[
\begin{align*}
w & = [0.000, 0.000, 1.060, 0.652, -0.652, -0.342, 1.701, 0.000, -0.658, 0.000, \\
& \quad 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.060, -0.239, 0.668, 0.391, 1.299, \\
& \quad -0.821, 0.000, 0.158, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.065, \\
& \quad 0.408, -0.234, 0.234, 0.272, 1.620, 0.908, 0.000, 0.000, 0.000, 0.000, 0.000, \\
& \quad 0.000, 0.000, 0.000, -0.255, 0.592, 1.234, -0.293, -0.332, 0.000, 0.092, \\
& \quad 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]
\end{align*}
\]

\( b = 0.500 \)
J.2 Newark Prediction Vector

\(w\)-vector at divider point 41 for time period 0800

\[
w = \begin{bmatrix} 0.002, -0.368, 0.251, 0.034, 0.061, 0.000, -0.028, -0.054, 0.161, -0.001, \\ 0.000, 0.000, 0.000, 0.000, 0.002, -0.191, -0.103, 0.055, 0.002, -0.769, \\ 0.000, 0.242, 0.000, 0.001, 0.000, 0.000, 0.000, 0.012, -0.005, 0.121, \\ 0.257, 0.531, 0.049, -0.041, 1.402, -0.221, 0.000, -0.001, 0.000, 0.000, \\ 0.000, 0.003, 0.003, -0.086, -0.188, -0.508, 0.027, 0.143, -0.377, 0.098, \\ 0.000, 0.000, 0.000, 0.000, 0.000, -0.018, -0.010 \end{bmatrix}
\]

\(b = 2.176\)

\(w\)-vector at divider point 43 for time period 0800

\[
w = \begin{bmatrix} 0.032, -0.147, 0.221, 1.441, -0.027, 0.854, 1.039, 0.274, 0.000, -0.003, \\ -0.001, -0.003, -0.001, 0.023, 0.012, -0.124, 0.224, 0.880, 0.061, -1.714, \\ 0.420, 0.361, 0.000, 0.009, 0.001, 0.002, 0.000, -0.278, -0.048, 0.105, \\ 0.681, 1.025, 0.075, -0.437, 1.293, -0.446, 0.000, -0.002, 0.001, -0.001, \\ 0.002, 0.597, 0.012, -0.020, -0.125, -0.584, 0.062, 0.949, -0.657, 0.446, \\ -0.503, -0.002, -0.001, 0.000, -0.001, -0.251, -0.550 \end{bmatrix}
\]

\(b = 2.984\)
$w$-vector at divider point 41 for time period 1000

$$w = [0.002, -0.318, 0.213, 0.117, 0.012, 0.000, -0.074, -0.239, 0.388, -0.001, 0.000, 0.000, 0.002, 0.003, -0.222, -0.098, 0.389, 0.030, -0.564, 0.007, 0.376, 0.000, 0.000, 0.000, 0.000, 0.019, -0.007, 0.098, 0.242, 0.304, 0.006, 0.005, 1.141, -0.161, 0.000, 0.000, 0.000, 0.000, 0.000, -0.018, 0.004, -0.144, -0.152, -0.477, 0.006, 0.157, -0.911, 0.070, 0.000, 0.000, 0.000, 0.000, 0.000, -0.008, -0.019]$$

$$b = 2.558$$

$w$-vector at divider point 43 for time period 1000

$$w = [0.035, -0.132, 0.156, 0.908, 0.059, 1.241, 1.112, 0.261, 0.000, -0.004, -0.001, -0.003, 0.000, 0.024, 0.016, -0.098, 0.333, 1.192, 0.084, -1.975, 0.480, 0.447, 0.000, 0.009, 0.001, 0.002, 0.000, -0.201, -0.043, 0.108, 0.634, 0.916, 0.089, -0.550, 1.355, -0.400, 0.000, -0.001, 0.001, -0.002, 0.001, 0.467, 0.016, -0.097, -0.131, -0.732, -0.041, 0.796, -1.092, 0.440, -0.189, -0.002, -0.001, 0.000, -0.001, -0.244, -0.573]$$

$$b = 3.074$$
$w$-vector at divider point 41 for time period 1600

$$w = [0.000, -0.186, -0.127, 0.610, 0.000, -0.246, -0.186, -0.186, 0.644, 0.000, 0.000, 0.000, 0.000, 0.000, -0.314, 0.280, -0.169, -0.127, -2.534, -0.322, 0.186, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.237, 1.076, 1.254, 0.000, 0.186, 0.653, -0.229, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.364, -0.229, -0.364, 0.000, 0.000, -0.856, 0.229, 0.847, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]$$

$$b = 2.763$$

$w$-vector at divider point 43 for time period 1600

$$w = [0.039, -0.286, 0.094, -0.223, 0.122, -0.651, -0.317, 0.012, 0.000, -0.003, 0.000, 0.000, 0.000, 0.033, -0.006, -0.210, -0.127, 1.071, -0.221, -1.587, 0.486, 0.082, 0.000, 0.000, 0.009, 0.000, 0.001, -0.001, -0.025, 0.002, 0.007, 1.087, 1.207, -0.072, -0.168, 0.486, -0.613, 0.000, -0.001, 0.001, 0.000, 0.003, 0.215, 0.030, -0.251, 0.362, -0.521, -0.034, 0.054, -2.404, 0.424, 0.359, -0.002, -0.001, 0.000, -0.002, -0.141, -0.233]$$

$$b = 3.972$$
\textit{w}-vector at divider point 41 for time period 1900

\[ w = [0.000, -0.176, -0.118, 0.359, 0.000, -0.353, 0.046, -0.176, 0.712, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, -0.353, -0.059, -0.346, 0.000, -2.059, \\
-0.353, 0.235, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.294, \\
1.176, 1.405, 0.000, 0.176, 1.294, -0.235, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, -0.412, 0.000, -0.412, 0.000, 0.000, -1.000, 0.118, \\
0.771, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] \]

\[ b = 2.882 \]

\textit{w}-vector at divider point 43 for time period 1900

\[ w = [0.032, -0.234, 0.124, -0.212, -0.019, -0.589, -0.295, -0.046, 0.056, -0.003, \\
0.000, -0.001, 0.000, -0.020, -0.006, -0.160, -0.145, 0.579, -0.139, -1.296, \\
0.785, 0.288, 0.000, 0.007, 0.000, 0.000, 0.001, 0.000, -0.042, -0.007, 0.132, \\
0.854, 1.267, 0.134, 0.142, 0.785, -0.300, 0.000, -0.001, 0.001, 0.001, \\
0.002, 0.293, 0.036, -0.260, 0.611, -0.266, -0.074, -0.180, -2.183, 0.197, \\
0.361, -0.003, -0.001, 0.000, -0.003, -0.121, -0.327] \]

\[ b = 3.217 \]
J.3 O’Hare Prediction Vector

\( w \)-vector at divider point 90 for time period 0800

\[
\begin{align*}
w & = \begin{bmatrix}
0.000, & 0.000, & 0.000, & 1.000, & 0.000, & 0.000, & 0.000, & 0.000, & 2.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & -0.333, & 0.000, & -0.667, & 0.333, & 0.000, \\
1.000, & 1.333, & -1.333, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.333, & -0.333, & 0.000, & 0.000, & -0.667, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & -0.333, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000
\end{bmatrix} \\
\quad \\
b & = 1.000
\end{align*}
\]

\( w \)-vector at divider point 96 for time period 0800

\[
\begin{align*}
w & = \begin{bmatrix}
0.000, & 0.000, & 0.000, & 2.000, & 0.000, & 2.000, & 0.000, & 0.000, & 2.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
2.000, & 2.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000
\end{bmatrix} \\
\quad \\
b & = -1.000
\end{align*}
\]
$w$-vector at divider point 90 for time period 1000

\[
w = \begin{bmatrix}
0.000, 0.000, 0.000, 1.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, -0.500, 0.000, -0.500, 0.000, 0.000,
0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
\end{bmatrix}
\]

\[
b = 2.000
\]

$w$-vector at divider point 96 for time period 1000

\[
w = \begin{bmatrix}
0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, -0.400, 0.800, -0.800, 0.400, 0.000,
0.000, 0.800, 0.400, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, -0.400, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
\end{bmatrix}
\]

\[
b = 1.400
\]
$w$-vector at divider point 90 for time period 1200

$$ w = [0.008, -0.048, -0.011, 1.298, -0.143, -0.082, 0.289, -0.143, 0.614, 0.000,
0.000, 0.000, 0.000, 0.000, 0.003, -0.081, 0.988, 0.046, 0.169, -0.808,
0.000, 0.153, 1.190, 0.002, 0.000, 0.000, -0.001, -0.097, -0.002, -0.321,
0.203, 0.162, -0.244, -0.418, 0.000, -0.239, 0.000, -0.002, 0.000, 0.000,
0.001, 0.051, 0.009, 0.198, 0.033, 0.017, 0.140, 0.278, 1.595, 0.184,
-1.617, 0.000, 0.000, 0.000, 0.000, -0.110, 0.046] $$

$$ b = 0.096 $$

$w$-vector at divider point 96 for time period 1200

$$ w = [0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000,
0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] $$

$$ b = -1.000 $$
$w$-vector at divider point 90 for time period 1900

$$w = [0.000, -0.028, 0.072, 0.065, 0.229, -0.047, -0.071, -0.016, 0.939, 0.000, 0.000, 0.000, -0.037, 0.001, -0.008, 0.101, 0.112, -0.105, -0.363, -1.560, -0.015, 0.939, 0.001, 0.000, 0.000, -0.001, -0.118, 0.005, -0.248, 0.486, -0.531, -0.317, 0.160, 0.000, -0.684, 1.729, -0.001, 0.000, 0.000, 0.001, 0.049, 0.004, -0.206, -0.150, 0.148, 0.329, 0.087, 0.685, 0.806, 0.016, 0.000, 0.000, 0.000, 0.000, -0.029, 0.034]$$

$$b = 1.616$$

$w$-vector at divider point 96 for time period 1900

$$w = [0.001, -0.002, 0.237, 0.012, 1.165, 0.760, 0.003, 0.005, 1.114, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.001, 0.000, -0.005, 0.000, -0.007, 0.875, 0.370, -0.343, -0.361, -0.030, -0.019, 1.114, 0.000, 0.000, 0.000, 0.000, 0.000, -0.015, 0.000, -0.162, 0.453, -0.550, 0.201, -0.398, -0.706, -0.971, 1.700, 0.000, 0.000, 0.000, 0.000, 0.014, 0.002, -0.243, -0.132, 0.115, 0.155, 0.593, 0.739, 1.059, 0.020, 0.000, 0.000, 0.000, 0.000, -0.012, 0.002]$$

$$b = 1.443$$
J.4 Atlanta Prediction Vector

\( w \)-vector at divider point 115 for time period 0800

\[
\begin{align*}
w &= [0.010, -0.383, 0.516, -1.257, 0.416, 0.567, 0.405, 0.266, 0.000, 0.005, \\
&\quad -0.001, 0.000, -0.001, 0.125, 0.011, 0.003, -0.480, 0.342, 0.643, 0.539, \\
&\quad -0.578, 0.439, -0.461, 0.003, -0.001, -0.001, 0.000, -0.040, 0.005, -0.047, \\
&\quad 0.471, 0.000, -0.141, -0.735, 1.599, 0.272, 0.000, -0.001, -0.001, 0.002, \\
&\quad 0.000, 0.026, -0.001, 0.041, 0.042, -0.234, 0.117, 0.481, 0.000, 0.058, \\
&\quad 0.000, 0.001, 0.001, -0.001, -0.006, -0.146, -0.089] \\
b &= 1.363
\end{align*}
\]

\( w \)-vector at divider point 124 for time period 0800

\[
\begin{align*}
w &= [0.085, -0.297, 0.498, 0.000, -0.226, 0.978, 2.227, 0.831, 0.000, 0.005, \\
&\quad -0.001, 0.001, -0.002, 0.095, 0.062, -0.010, -0.166, 0.000, 0.522, 0.368, \\
&\quad -1.919, 0.235, 0.000, 0.007, 0.000, 0.000, -0.002, -0.290, 0.018, 0.059, \\
&\quad 0.202, 0.000, -0.732, -0.828, 0.296, 0.364, 0.000, -0.001, 0.000, 0.001, \\
&\quad 0.000, -0.128, -0.014, -0.068, 0.224, 1.261, 0.368, 0.842, 0.296, 0.120, \\
&\quad 0.000, -0.001, 0.001, -0.001, 0.004, 0.039, -0.278] \\
b &= 1.585
\end{align*}
\]
\( w \)-vector at divider point 115 for time period 1100

\[
w = \begin{bmatrix}
0.002, -0.312, 0.450, -1.236, 0.152, 2.124, 0.441, 0.133, 0.043, 0.001,
0.000, 0.000, -0.001, 0.061, 0.013, -0.046, -0.084, 0.404, 0.838, 0.221,
-0.488, 0.423, -0.102, 0.007, -0.002, 0.000, 0.001, -0.182, -0.002, 0.086,
1.070, 0.302, -0.824, -0.571, 1.639, 0.464, 0.000, -0.002, 0.000, 0.001,
0.000, 0.256, 0.001, 0.026, -0.083, -0.202, 0.262, 0.125, 0.000, 0.167,
0.000, 0.002, 0.000, -0.001, -0.004, -0.210, -0.060
\end{bmatrix}
\]

\( b = 0.539 \)

\( w \)-vector at divider point 124 for time period 1100

\[
w = \begin{bmatrix}
0.055, -0.163, 0.901, 0.004, -0.638, 1.256, 1.451, 1.074, 0.000, 0.008,
0.000, 0.002, -0.002, -0.016, 0.066, -0.164, -0.429, 0.000, 0.335, 0.344,
-3.040, 0.053, 0.000, 0.006, -0.002, 0.002, -0.004, -0.229, -0.003, -0.138,
-0.051, 0.000, 0.189, -0.249, 0.364, -0.281, 0.000, -0.001, 0.001, 0.003,
0.002, 0.303, 0.012, -0.023, 0.530, 1.484, 0.191, 0.080, 0.364, 0.327,
0.000, 0.000, -0.001, -0.002, -0.001, -0.396, -0.148
\end{bmatrix}
\]

\( b = 2.257 \)
$w$-vector at divider point 115 for time period 1600

$$w = [-0.001, -0.156, 0.058, 1.724, 0.222, 0.000, 0.751, 0.361, 0.523, 0.002,$$

$$0.001, 0.000, 0.048, 0.010, -0.113, 0.431, -1.126, 0.075, 0.014,$$

$$-1.291, 0.070, -1.462, 0.008, -0.001, -0.001, 0.000, 0.039, -0.008, 0.075,$$

$$0.918, 0.305, -0.288, -0.200, 1.701, -0.022, 0.469, -0.002, -0.001, 0.001,$$

$$0.001, -0.097, 0.013, -0.069, -0.355, 0.305, 0.442, 0.887, -0.535, 0.577,$$

$$0.469, 0.000, -0.001, -0.001, -0.004, -0.237, -0.092]$$

$$b = 0.781$$

$w$-vector at divider point 124 for time period 1600

$$w = [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,$$

$$0.000, 0.000, -0.008, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,$$

$$-2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,$$

$$0.000, 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000,$$

$$0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000,$$

$$0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]$$

$$b = 1.000$$

220
**w-vector at divider point 115 for time period 2000**

\[ w = \begin{bmatrix} 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000 \end{bmatrix} \]

\[ b = -1.000 \]

**w-vector at divider point 124 for time period 2000**

\[ w = \begin{bmatrix} 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, -0.008, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000 \end{bmatrix} \]

\[ b = 1.000 \]
J.5  LaGuardia Prediction Vector

\( w \)-vector at divider point 36 for time period 0700

\[
w = [0.075, -0.067, 0.031, 0.865, -0.159, 1.921, 0.825, -0.087, 0.000, 0.002, 0.000, 0.000, 0.000, 0.283, -0.008, 0.088, -0.399, -0.523, 0.659, -0.457, 0.000, 0.553, 0.000, -0.003, 0.000, 0.002, 0.000, -0.186, -0.007, -0.159, -0.438, 0.253, 0.030, 0.004, 0.018, 0.495, 0.000, 0.001, 0.001, -0.001, 0.002, -0.121, 0.002, 0.057, -0.549, -0.606, 0.193, 0.358, -0.413, 0.470, 0.993, -0.003, -0.002, -0.001, -0.002, 0.125, -0.273]
\]

\( b = 0.358 \)

\( w \)-vector at divider point 39 for time period 0700

\[
w = [0.032, -0.144, 0.356, 0.666, -0.113, 0.679, -0.240, -0.178, 0.204, 0.000, 0.001, -0.001, -0.001, 0.126, -0.010, 0.059, 0.468, -0.678, 0.053, -0.047, 0.204, -0.094, 0.000, -0.001, 0.001, 0.001, -0.003, -0.275, 0.004, 0.011, -0.191, 1.189, -0.460, 0.289, -0.431, -0.117, 0.000, 0.000, -0.002, -0.001, 0.003, 0.067, 0.006, -0.112, -0.406, -1.037, 0.234, 0.118, -0.007, 0.802, 0.424, -0.003, 0.001, 0.001, -0.002, -0.048, -0.674]
\]

\( b = 4.372 \)
$w$-vector at divider point 36 for time period 1000

\[
w = \begin{bmatrix}
0.050, -0.072, 0.084, 0.295, -0.279, 1.522, 1.186, 0.124, 0.000, 0.003, \\
0.000, 0.000, 0.001, 0.250, 0.016, -0.101, -0.704, -0.555, 0.692, 0.029, \\
0.000, 0.659, 0.000, -0.005, 0.000, 0.002, 0.001, -0.207, -0.016, 0.013, \\
0.402, 1.272, -0.327, -0.561, -1.255, -0.251, 0.000, 0.000, 0.001, -0.001, \\
0.002, -0.289, 0.013, 0.061, -0.657, -0.898, 0.091, 0.391, 0.759, 0.806, \\
0.694, -0.002, -0.002, 0.000, -0.003, 0.183, -0.324
\end{bmatrix}
\]

\[b = 0.628\]

$w$-vector at divider point 39 for time period 1000

\[
w = \begin{bmatrix}
0.023, -0.077, 0.601, 0.627, -0.367, -0.224, 0.295, -0.075, 0.000, 0.000, \\
0.000, -0.001, 0.000, 0.090, -0.001, -0.051, 0.199, -0.828, 0.155, 0.426, \\
0.559, 0.324, 0.000, -0.001, 0.001, 0.000, -0.001, -0.182, 0.005, 0.129, \\
0.155, 1.592, -0.348, 0.027, -1.785, -0.232, 0.000, -0.003, -0.002, -0.001, \\
0.001, 0.157, 0.006, -0.120, -0.547, -1.074, 0.212, 0.411, 0.078, 0.446, \\
0.078, 0.000, 0.000, 0.000, -0.002, -0.252, -0.646
\end{bmatrix}
\]

\[b = 3.862\]
$w$-vector at divider point 36 for time period 1400

$$w = \begin{bmatrix} 0.029, -0.134, 0.115, 0.509, -0.454, 1.354, 0.161, -0.073, 0.265, 0.005, \\
0.001, 0.000, 0.000, 0.152, 0.014, -0.104, -0.602, -0.677, 0.718, 0.000, \\
-1.094, 0.519, 0.265, -0.003, 0.000, 0.001, 0.000, 0.034, -0.004, 0.101, \\
0.775, 1.560, -0.445, 0.060, 0.874, -0.241, 0.000, -0.003, 0.001, 0.000, \\
0.000, -0.227, 0.021, -0.020, -0.566, -0.472, 0.149, 0.439, 0.488, 0.813, \\
0.000, 0.000, -0.001, -0.002, 0.000, 0.094, -0.214 \end{bmatrix}$$

$$b = 0.370$$

$w$-vector at divider point 39 for time period 1400

$$w = \begin{bmatrix} 0.017, -0.138, 0.217, 0.762, 0.230, 0.020, -0.339, -0.160, 0.000, 0.000, \\
0.000, 0.001, 0.000, -0.005, 0.003, -0.151, -0.246, -0.618, 0.430, 0.000, \\
0.591, 0.087, 0.000, -0.002, 0.001, -0.001, -0.001, 0.006, 0.021, 0.268, \\
0.388, 0.556, -0.349, -0.170, -0.881, -0.060, 0.000, -0.001, -0.002, 0.000, \\
0.000, -0.127, -0.005, -0.169, -0.337, -0.720, -0.054, 0.541, 3.119, 0.547, \\
0.000, 0.000, -0.002, -0.001, -0.004, -0.191, -0.427 \end{bmatrix}$$

$$b = 3.429$$
w-vector at divider point 36 for time period 1800

\[
w = \begin{bmatrix} 0.010, & -0.093, & 0.297, & -0.132, & -0.350, & 1.647, & 0.432, & -0.147, & 0.506, & 0.005, \\
0.000, & 0.000, & 0.000, & 0.074, & 0.010, & -0.165, & -0.554, & 0.055, & 0.897, & 0.080, \\
-1.366, & 0.490, & 0.169, & -0.004, & 0.000, & 0.001, & 0.000, & -0.001, & -0.012, & -0.046, \\
0.883, & 0.364, & -0.308, & 0.074, & 1.021, & -0.126, & 0.000, & -0.001, & 0.000, & 0.000, \\
0.000, & -0.106, & 0.024, & 0.033, & -0.211, & 0.163, & -0.092, & 0.170, & -0.044, & 0.246, \\
0.370, & -0.001, & 0.000, & -0.001, & 0.000, & -0.029, & -0.102 \end{bmatrix}
\]

\[b = 0.749\]

w-vector at divider point 39 for time period 1800

\[
w = \begin{bmatrix} 0.008, & -0.095, & -0.011, & 0.246, & 0.107, & 0.078, & -0.401, & 0.034, & 0.026, & 0.001, \\
0.000, & 0.001, & 0.000, & 0.047, & 0.003, & -0.080, & -0.286, & -0.554, & 0.201, & 0.239, \\
0.083, & 0.193, & 0.000, & -0.002, & 0.001, & -0.001, & -0.003, & 0.024, & 0.005, & 0.073, \\
0.449, & 0.070, & -0.338, & -0.292, & 0.137, & -0.144, & 0.000, & -0.001, & -0.002, & 0.000, \\
-0.002, & -0.076, & 0.010, & 0.043, & 0.224, & 0.211, & -0.111, & 0.041, & 1.061, & 0.097, \\
0.000, & 0.001, & -0.001, & 0.000, & -0.005, & -0.137, & -0.218 \end{bmatrix}
\]

\[b = 1.857\]
J.6 JFK Prediction Vector

\( w \)-vector at divider point 49 for time period 1600

\[
\begin{align*}
w &= [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000] \\
b &= -1.000
\end{align*}
\]

\( w \)-vector at divider point 53 for time period 1600

\[
\begin{align*}
w &= [0.082, -0.177, -0.188, 1.314, 0.275, -0.467, -0.592, -0.164, 0.001, 0.002, -0.001, 0.000, -0.001, 0.027, 0.012, -0.067, 0.480, 0.410, 0.071, -0.661, 1.077, -0.279, 0.000, -0.002, 0.001, 0.000, -0.001, 0.587, 0.018, 0.047, -0.336, 0.410, 0.426, 0.294, 0.239, 0.677, 0.000, 0.003, -0.001, -0.001, -0.003, -0.611, -0.018, 0.065, 0.723, 1.424, -0.364, -0.091, 0.000, -0.284, 0.000, -0.002, 0.001, 0.000, 0.000, -0.786, -0.003] \\
b &= 1.007
\end{align*}
\]
J.7 Reagan National Prediction Vector

\( w \)-vector at divider point 31 for time period 0800

\[
w = \begin{bmatrix}
0.003, -0.014, 0.867, 0.008, -0.422, -1.011, 0.020, 0.010, 0.546, -0.001, \\
0.000, 0.000, 0.000, -0.040, -0.004, -0.370, -0.297, 1.819, -0.108, 0.160, \\
0.065, 0.477, 0.269, -0.002, 0.000, 0.000, 0.000, 0.008, 0.003, 0.028, \\
-0.106, -1.313, 0.163, 0.021, 0.442, -0.138, 0.269, 0.001, 0.000, 0.000, \\
0.000, -0.079, 0.000, 0.012, 0.169, -0.099, -0.220, -0.050, 0.376, -0.589, \\
0.108, -0.001, 0.000, 0.000, 0.000, 0.071, -0.010
\end{bmatrix}
\]

\[
b = 1.051
\]

\( w \)-vector at divider point 35 for time period 0800

\[
w = \begin{bmatrix}
0.004, -0.056, 1.273, -0.003, -0.099, -1.680, 0.303, 0.029, 0.000, 0.000, \\
0.000, 0.000, 0.000, -0.063, -0.005, -0.327, -0.780, 1.395, 0.204, 0.605, \\
0.000, 0.717, 0.263, 0.000, 0.000, 0.000, 0.000, 0.000, 0.004, 0.005, 0.000, \\
-0.020, -0.916, 0.412, -0.183, 0.184, 0.093, 0.263, -0.001, -0.001, 0.000, \\
-0.001, -0.070, 0.009, 0.052, 0.185, 0.173, -0.054, 0.152, 0.184, -0.144, \\
0.263, 0.000, 0.000, 0.000, 0.001, 0.008, 0.071
\end{bmatrix}
\]

\[
b = 0.513
\]
$w$-vector at divider point 31 for time period 1300

$$w = [0.004, -0.018, 0.830, 0.027, -0.632, -0.992, 0.057, 0.041, 0.314, 0.000, -0.001, 0.000, 0.000, -0.054, -0.001, -0.332, -0.255, 1.464, -0.170, 0.182, 0.000, 0.092, 0.200, -0.001, 0.000, 0.000, 0.000, -0.030, 0.003, -0.030, -0.089, -1.349, 0.246, 0.026, 0.514, 0.016, 0.200, -0.001, -0.001, 0.000, -0.004, -0.066, -0.003, 0.021, 0.142, 0.223, -0.111, -0.097, 0.514, -0.101, 0.200, -0.001, 0.000, 0.000, 0.000, 0.072, -0.021]$$

$$b = \ 1.164$$

$w$-vector at divider point 35 for time period 1300

$$w = [0.000, 0.000, 1.429, 0.000, -0.571, -1.429, 0.571, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.286, -0.857, 1.143, 0.286, 0.571, 0.095, 0.857, 0.286, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.857, 0.571, 0.000, 0.191, 0.000, 0.571, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]$$

$$b = \ 0.714$$
$w$-vector at divider point 31 for time period 1600

\[
w = [0.004, -0.017, 0.476, 0.000, -0.352, -0.552, 0.267, 0.161, 0.989, 0.001, \\
0.000, 0.000, 0.000, -0.177, -0.002, -0.273, 0.235, 0.482, -0.543, -0.382, \\
0.000, 0.134, 1.251, -0.001, 0.000, 0.000, 0.000, -0.077, 0.011, 0.085, \\
0.086, 0.405, 0.210, -0.083, 0.426, 0.037, -0.003, 0.000, -0.001, 0.000, \\
-0.011, -0.100, -0.008, -0.039, 0.186, 0.146, -0.100, -0.095, 0.426, -0.101, \\
0.000, -0.002, -0.001, 0.000, 0.001, 0.112, -0.023]
\]

$\mathbf{b} = 0.360$

$w$-vector at divider point 35 for time period 1600

\[
w = [0.000, 0.000, 0.000, 0.000, 0.000, -2.000, 0.000, 0.000, 2.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.667, 2.000, 0.667, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.667, 0.000, \\
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.667, 0.000, \\
0.667, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]
\]

$\mathbf{b} = -1.000$
$w$-vector at divider point 31 for time period 1800

$$w = \begin{bmatrix} 0.004, -0.018, 0.433, -0.217, -0.333, -0.418, 0.043, 0.135, 0.996, 0.001, \\
0.000, 0.000, 0.001, -0.164, -0.003, -0.272, 0.200, 0.898, -0.500, -0.442, \\
0.000, 0.065, 1.334, -0.001, 0.000, 0.000, 0.000, -0.015, 0.009, 0.031, \\
0.120, 0.019, 0.137, -0.177, 0.351, -0.062, 0.035, 0.000, -0.001, 0.000, \\
-0.011, -0.113, -0.007, 0.005, 0.175, 0.289, -0.075, -0.038, 0.351, -0.024, \\
0.000, -0.002, -0.001, 0.000, 0.000, 0.105, -0.020 \end{bmatrix}$$

$$b = 0.429$$

$w$-vector at divider point 35 for time period 1800

$$w = \begin{bmatrix} 0.007, -0.023, 1.142, 0.000, -0.901, -1.681, 0.417, 0.203, 1.854, 0.001, \\
0.000, 0.000, 0.000, -0.092, 0.000, -0.117, -0.245, -0.141, -0.043, 0.507, \\
-0.022, 0.967, 0.548, 0.001, 0.000, 0.000, 0.000, 0.063, -0.007, 0.016, \\
-0.142, 0.384, 0.487, -0.178, 0.233, 0.080, 0.548, -0.001, 0.000, 0.001, \\
-0.003, -0.155, 0.013, -0.044, 0.343, 0.595, 0.221, 0.090, 0.255, -0.017, \\
0.548, 0.000, 0.000, 0.000, 0.003, -0.031, 0.190 \end{bmatrix}$$

$$b = -1.054$$
J.8 Dulles Prediction Vector

\(w\)-vector at divider point under 80 for time period 0700

\[
\begin{align*}
\boldsymbol{w} & = [0.000, -0.194, 1.051, 0.264, -0.523, -2.161, 1.032, -0.194, 0.003, 0.000, \\
& 0.000, 0.000, 0.000, 0.000, 0.090, -0.290, 0.510, 0.381, -0.039, \\
& -0.309, 1.316, 1.355, 0.000, 0.000, 0.000, 0.000, 0.000, -0.065, \\
& -0.065, -0.129, 0.000, -0.381, 1.213, 0.091, 0.391, 0.000, 0.000, 0.000, \\
& 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, -0.013, 0.000, -1.464, 0.000, \\
& 0.390, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]
\end{align*}
\]

\[
\begin{align*}
\boldsymbol{b} & = 0.006
\end{align*}
\]

\(w\)-vector at divider point over 80 for time period 0700

\[
\begin{align*}
\boldsymbol{w} & = [-0.036, 0.062, -0.014, -0.330, 0.076, -0.577, -0.308, -0.256, 0.204, 0.000, \\
& 0.002, 0.001, 0.003, -0.264, -0.034, 0.120, 0.033, 0.684, -0.209, 0.663, \\
& 0.000, 0.018, -0.161, -0.001, -0.001, 0.002, 0.002, -0.051, 0.018, -0.127, \\
& 0.118, -0.630, 0.067, -0.291, 0.000, -0.379, 0.000, 0.001, 0.001, 0.002, \\
& 0.001, -0.070, 0.015, 0.047, 0.046, -0.062, -0.091, -0.011, 0.000, 0.098, \\
& 0.000, 0.002, 0.000, 0.000, 0.000, 0.002, 0.031, -0.050]
\end{align*}
\]

\[
\begin{align*}
\boldsymbol{b} & = -1.068
\end{align*}
\]
$w$-vector at divider point under 80 for time period 1100

$$w = \begin{bmatrix}
0.000, & 0.000, & 1.000, & 1.000, & -1.000, & -1.000, & 1.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 1.000, & 0.000, \\
0.000, & 1.000, & 1.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 2.000, & 0.000, & 0.000, & 0.000, & 0.000, & 1.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & -1.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000 \\
\end{bmatrix}$$

$b = -1.000$

$w$-vector at divider point over 80 for time period 1100

$$w = \begin{bmatrix}
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, & 0.000, \\
\end{bmatrix}$$

$b = -1.000$
\( w \)-vector at divider point under 80 for time period 1500

\[
 w = \begin{bmatrix}
 0.000, 0.000, 2.000, 0.000, 0.000, -2.000, 0.000, 0.000, 0.000, \\
 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, \\
 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \\
 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000 \\
\end{bmatrix}
\]

\( b = -1.000 \)

\( w \)-vector at divider point over 80 for time period 1500

\[
 w = \begin{bmatrix}
 -0.010, 0.034, -0.124, -0.695, -0.109, 0.000, -0.062, -0.204, -0.167, -0.002, \\
 -0.001, -0.001, 0.003, -0.136, -0.023, 0.004, 0.051, -0.254, -0.036, 0.027, \\
 0.075, -0.164, 0.000, 0.000, -0.001, 0.004, 0.002, -0.023, 0.000, -0.026, \\
 0.000, 0.000, -0.162, -0.447, 0.000, 0.047, 0.000, 0.000, 0.001, 0.002, \\
 0.002, -0.141, -0.009, 0.089, -0.053, -0.606, -0.019, -0.044, 0.000, -0.078, \\
 0.000, 0.003, 0.001, 0.000, 0.001, -0.080, -0.067 \\
\end{bmatrix}
\]

\( b = -0.664 \)
w-vector at divider point under 80 for time period 2000

\[ w = \begin{bmatrix} 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \ldots \\ 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.200, 2.000, 0.000, -2.000, \\ 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \ldots \\ 0.000, -2.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \ldots \\ 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 2.000, 0.000, 0.000, 0.000, 0.000, \ldots \\ 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, \ldots \end{bmatrix} \]

\[ b = -1.000 \]

w-vector at divider point over 80 for time period 2000

\[ w = \begin{bmatrix} 0.003, -0.018, 0.034, -0.865, -0.105, -0.603, 0.280, -0.357, -0.762, -0.002, \\ -0.001, -0.002, 0.004, -0.253, -0.042, 0.045, -0.060, -0.991, -0.090, 0.386, \\ -0.319, -0.062, 0.000, -0.002, 0.000, 0.004, 0.001, 0.090, 0.010, -0.119, \\ -0.291, -0.397, -0.090, -0.398, 0.455, -0.031, 0.000, 0.000, 0.000, 0.001, \\ 0.003, 0.116, -0.027, 0.112, 0.160, -0.747, 0.140, -0.263, 1.209, -0.116, \\ 0.000, 0.000, 0.001, 0.000, 0.002, -0.310, 0.076 \end{bmatrix} \]

\[ b = -0.543 \]
Bibliography
Bibliography


237


[34] Ground delay programs. Federal Aviation Administration, 2006.


LTC David A. Smith was born in Lansdale, Pennsylvania, grew up in New Jersey, Michigan, and graduated from Binghamton High School in Binghamton, New York. He is a 1990 graduate of the United States Military Academy with a Bachelor of Science in Aerospace Engineering. After graduation, he was commissioned as a Second Lieutenant in the United States Army and was initially stationed in Schweinfurt, Germany as an armor officer. After Germany, the army career sent him to Fort Knox, Kentucky and Fort Carson, Colorado. In May of 2000, he graduated from Rensselaer Polytechnic Institute with a Master of Science in Applied Mathematics. After graduation, he was assigned to the United States Military Academy where he taught calculus and differential equations. In 2003, he was assigned to work for the Army Basing Study which was responsible for the Army’s recommendations for the 2005 Base Realignment and Closure round. His next assignment is as the Chief of Assessment for the Multi-National Security Transition Corps, Iraq.